UNIVERSITÉ DE MONTRÉAL

FAST AND SEAMLESS MOBILITY MANAGEMENT IN IPV6-BASED NEXT-GENERATION WIRELESS NETWORKS

LI JUN ZHANG
DÉPARTEMENT DE GÉNIE INFORMATIQUE ET GÉNIE LOGICIEL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

THÈSE PRÉSENTÉE EN VUE DE L'OBTENTION
DU DIPLÔME DE PHILOSOPHIÆ DOCTOR
(GÉNIE INFORMATIQUE)
AOÛT 2008

Your file Votre référence
ISBN: 978-0-494-46123-5
Our file Notre référence
ISBN: 978-0-494-46123-5

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Cette thèse intitulée :

FAST AND SEAMLESS MOBILITY MANAGEMENT IN IPV6-BASED NEXT-GENERATION WIRELESS NETWORKS

présentée par : Li Jun ZHANG
en vue de l'obtention du diplôme de : Philosophiæ Doctor
a été dûment accepté par le jury constitué de :

Mme NICOLESCU Gabriela, Doct., présidente.
M. PIERRE Samuel, Ph.D., membre et directeur de recherche.
M. QUINTERO Alejandro, Doct., membre.
Mme CHERKAOUI Soumaya, Ph.D., membre.

To my friends, well wishers, teachers,
and my family: Isaac and Caleb.

ACKNOWLEDGEMENTS

This research work presented in this thesis was done during my staying at Mobile Computing and Networking Research Laboratory (LARIM), Department of Computer Engineering and Software Engineering, École Polytechnique de Montréal, Montreal, Quebec, Canada.

To my advisor, Professor Samuel Pierre, the director of LARIM, I would like to express my gratitude for offering me the challenge and financial support to pursue my Ph.D. studies in the field of mobility management, integrated architecture design, optimization, and system engineering, for his fruitful advice and guidance throughout my thesis, and for his encouragement and stimulation which helped me drive my research work to a successful conclusion.

My thanks go to the experts of Ericsson Research Canada, especially Mr. Laurent Marchand and Mr. Yves Lemieux, for their helpful advice.

My thanks also go to all my colleagues at the LARIM for providing me a pleasant and stimulating environment. Among the numerous colleagues, Dr. Desire Oulai, Mr. Paul Vital Hiol Mahop, Mrs. Sabine Kébreau and Mrs. Betty Momplaisir deserve a special mention for many invaluable suggestions.

I would also like to thank Professors Alejandro Quintero, Steven Chamberland and Abderrahim Benslimane, the jury members of my Ph.D written and oral exams, who carefully reviewed my thesis proposal and provided me with numerous valuable comments.

RÉSUMÉ

Jusqu'à présent, avec les progrès rapides dans les réseaux sans fil, les utilisateurs mobiles ont été capables de bénéficier de systèmes sans fil disparates, tels que les réseaux sans fil personnels (WPANs) (par exemple Bluetooth), les réseaux locaux sans fil (WLAN), les réseaux métropolitains sans fil (WMANs), les réseaux sans fil étendus (WWANs) comme la troisième génération (3G) réseaux cellulaires, etc.

Par ailleurs, la prochaine génération de réseaux sans fil (NGWNs) devrait intégrer les différents réseaux ou systèmes sans fil et posséder une infrastructure tout-IP soutenant la mobilité entre les technologies d'accès hétérogènes. L'un des principaux défis pour la recherche de ces réseaux de prochaine génération est de concevoir des régimes efficaces de gestion de la mobilité qui permettent de s'acquitter de nœuds mobiles d'itinérance entre les différentes technologies d'accès.

Pour réaliser efficacement la gestion de la mobilité, l'Internet Engineering Task Force (IETF) a proposé plusieurs solutions de gestion de la mobilité, telles que l'appui à la mobilité IP version 4 (MIPv4), le protocole mobile IPv6 (MIPv6), hiérarchique mobile IPv6 (HMIPv6), relèves rapides pour mobile IPv6 (FMIPv6) et relève rapide pour hiérarchique mobile IPv6 (F-HMIPv6), proxy mobile IPv6 (PMIPv6), etc. Outre les activités de recherche dans les organismes de normalisation, un certain nombre de protocoles de gestion de la mobilité ont été proposés dans la littérature. Par exemple, le protocole de la gestion de mobilité intra-domaine (IDMP), IP cellulaire, le protocole de relèves conscients des infrastructures d'accès sans fil à Internet (HAWAI), etc. Cependant, aucun de ces protocoles ne constitue une solution idéale pour la mobilité sans coupure avec la provision de la qualité de service.

L'objectif de cette recherche est de concevoir de nouveaux protocoles qui soutiennent la gestion de mobilité rapide et sans coupure dans les réseaux sans fil de prochaine génération. De tels protocoles doivent présenter des caractéristiques comme un temps de relève réduit, une signalisation réduite, moins de taux de perte de paquets, moins de session d'interruption.

Plus précisément, dans cette thèse, nous proposons un nouveau Routeur d'Accès Tunneling Protocole (ARTP), qui permet aux routeurs d'accès (ARs) d'établir les tunnels sécurisés bidirectionnels (BSTs) avec leurs voisins. À partir de l'ARTP, nous analysons les architectures existantes intégrées dans les réseaux sans fil de prochaine

génération et nous proposons une nouvelle architecture. Nous analysons ensuite les protocoles de gestion de la mobilité qui sont proposés par les groupes de travail de l'IETF. Puis, nous proposons des solutions optimisées de relèves sans coupure pour les réseaux basés sur MIPv6. Ces solutions permettent aux noeuds mobiles d'exploiter leurs précédentes adresses IP dans les nouveaux réseaux visités, à condition que le nouveau réseau d'accès et celui qui le précède aient un accord de niveau de service avant les relèves actuelles. Cet accord permet aux utilisateurs mobiles d'utiliser les tunnels bidirectionnels pré-configurés, en particulier, pour ceux qui font la relève avec des sessions multimédias en cours. Comme les utilisateurs mobiles emploient les tunnels bidirectionnels pré-configurés au cours de la relève, la signalisation utilisée pour établir ces tunnels est retirée de l'ensemble du processus de relève. Cela réduit le temps de relève. En outre, la latence relative à la détection d'adresse dupliquée (DAD) est complètement éliminée du processus de relève. Cela minimise le temps de relève. Comme SMIPv6 permet aux noeuds mobiles d'utiliser leurs anciennes adresses IP (PCoA) immédiatement après avoir joint la nouvelle liaison, alors, les contextes d'informations des noeuds mobiles sont conservés intacts dans leur routeurs d'accès précédents (PARs). Par conséquent, les délais qui concernent le transfert de contextes sont également éliminés complètement de la latence de relève.

Pour évaluer l'efficacité de nos approches, nous utilisons des modèles analytiques pour étudier les performances de la relève et l'emplacement de gestion en comparant ces performances avec MIPv6, HMIPv6, FMIPv6 et F-HMIPv6. En outre, des simulations sont également réalisées pour les procédures de la relève en vue d'analyser les performances en termes de temps de relèves, taux de perte de paquets. Dans cette thèse, nous développons aussi une nouvelle approche pour la relève rapide de la couche deux (couche de MAC) dans des réseaux intégrés le protocole de MIPv6 et WLAN. Ensuite, nous utilisons des simulations pour évaluer la performance de la relève.

ABSTRACT

So far, with the rapid progress in wireless networking and mobile computing, mobile users have been capable to benefit from disparate wireless systems, such as wireless personal area networks (WPANs) (e.g. Bluetooth), wireless local area networks (WLANs), wireless metropolitan area networks (WMANs), wireless wide area networks (WWANs) like third generation (3G) cellular networks, etc.

On the other hand, next generation wireless networks (NGWNs) are expected to integrate existing various wireless networks/systems and to present an all-IP-based infrastructure to support mobility among heterogeneous radio access technologies. One of the major research challenges of such networks is to design effective mobility management schemes that enable mobile nodes (MNs) to perform roaming across various access technologies.

To realize efficient mobility management, the Internet Engineering Task Force (IETF) has proposed several mobility management solutions, such as mobile IPv4 (MIPv4), mobile IPv6 (MIPv6), hierarchical mobile IPv6 (HMIPv6), fast handovers for mobile IPv6 (FMIPv6), fast handover for hierarchical mobile IPv6 (F-HMIPv6), proxy mobile IPv6 (PMIPv6), etc. Besides the research activity within standardization bodies such as IETF, 3G partnership project initiatives, i.e. 3GPP and 3GPP2, a number of mobility management protocols have been proposed in the literature. For example, intra-domain mobility management protocol (IDMP), cellular IP, handoff-aware wireless access internet infrastructure (HAWAII) protocol, etc. However, none of these protocols provides a perfect solution for seamless mobility with quality of service (QoS) provisioning.

The objective of this research is to conceive new protocols that support fast and seamless mobility management in IPv6-based next-generation wireless networks. Such schemes should present the features such as optimal handoff latency, lower signalling overhead, less packet loss rate, user imperceptible interruption for sessions in progress.

More specifically, in this thesis, we initially propose a new access router tunnelling protocol (ARTP), which allows access routers (ARs) to establish bidirectional secure tunnels (BSTs) with their neighbors. Followed by the ARTP, we analyze existing integrated architectures for next-generation wireless networks and propose a new integrated architecture for such interworking systems. And then, we propose new

seamless handoff schemes (SMIPv6) for the integrating architecture. Such schemes allow mobile nodes to utilize their previous valid IP addresses, called previous care-of addresses (PCoAs) in new visiting networks, on condition that the visiting access network (AN) and the previous AN have made a service level agreement (SLA), within which specific MNs are allowed to exploit pre-configured bidirectional tunnels during handoff for their ongoing multimedia sessions. Since mobile users employ pre-established bidirectional tunnels during handoff, the signalling used for establishing such tunnels is removed from the overall handoff process. Thus results in reduced handoff delays. In addition, the latencies pertaining to the duplicate address detection (DAD) process are eliminated completely from the handoff process. This further minimizes the handoff delay. Moreover, since SMIPv6 schemes enable MNs to use their PCoAs immediately after attached to the new link, MNs' context information is kept intact at their previous ARs (PARs). Hence, the delays with respect to the context transfer process are also totally eliminated from the handoff latency. To evaluate the efficiency of our proposed SMIPv6 schemes, we use analytical models to investigate the impact of various wireless system parameters on the performance of the handoff process, and compare the obtained results with MIPv6, HMIPv6, FMIPv6 and F-HMIPv6. Moreover, simulations are carried out using OPNET Modeler v.12.0 to analyze the handoff performance in terms of end-to-end delay during handoff, packet drop rate, control traffic sent during handoff, etc. In addition, this thesis also develops a new fast MAC layer handoff scheme for a MIPv6/WLANs environment, and simulations are executed using the simulator SimulX to evaluate the handoff performance.

CONDENSÉ EN FRANÇAIS

Chapitre 1. Introduction

Avec les progrès rapides accomplis dans la technologie sans fil, le nombre d'utilisateurs mobiles a incroyablement augmenté ces dernières années. Comme de plus en plus d'utilisateurs sans fil représentent une majorité de la population de l'Internet aujourd'hui, la gestion de la mobilité devient un des principaux défis de la recherche dans la conception des systèmes sans fil.

1.1 Les motivations et les défis de recherche

Récemment, la conception de protocoles pour soutenir la mobilité aisée (ou sans coupure) tout en garantissant la qualité de service devient un sujet brûlant dans les réseaux sans fil hétérogènes tout-en-IP de la prochaine génération. Dans ces réseaux, les utilisateurs mobiles doivent disposer de la capacité d'itinérance commodément et sans coupure entre des différents opérateurs à travers différentes technologies d'accès.

En outre, la nouvelle génération de réseaux sans fil (NGWNs) a tendance à appuyer des services à valeur ajoutée (VAS). Cela impose une nouvelle complexité dans la conception des systèmes sans fil, parce que ces genres de services amènent des nouvelles exigences de qualité de service dans les réseaux sans fil de la prochaine génération. Aujourd'hui, le défi principal de la gestion de mobilité dans les réseaux sans fil est de considérer deux requis : 1) Minimiser le temps d'une relève de sorte que les utilisateurs mobiles ne peuvent pas constater le changement de leur réseau attaché. 2) Réduire le nombre de pertes de paquets autant que possible lors de la relève.

1.2 Les contributions et les originalités

Les principales contributions et originalités de cette thèse sont les suivantes :

- Cette thèse propose un nouveau protocole qui permet la tunnelisation entre deux routeurs d'accès adjacents. Ce protocole permet aussi l'établissement d'un

tunnel avec un ensemble de caractéristiques minimales entre deux routeurs (ou nœuds) dans un réseau. La nouveauté de ce protocole est d'ajouter un mécanisme à la fonction de tunnelisation traditionelle qui garantit une certaine qualité de service et un niveau de sécurité.La nouveauté de ce protocole est de définir un mécanisme qui permet aux nœuds (ou routeurs d'accès) d'établir les tunnels sécurisés bidirectionnels (BSTs) avec leurs voisins et de négocier les parameters de ces tunnels. La technique traditionnelle utilisée pour la tunnelisation entre deux nœuds ne peut pas garantir la qualité de service lors de l'utilisation des tunnels, le protocole de l'ARTP résolve ce problème par l'ajoute un mécanisme de négociation entre deux points de terminaison du tunnel. Par conséquence, certains paramètres concernant la qualité de service sont prédéfinis avant l'utilisation du tunnel. En outre, des points de terminaison du tunnel configurent des politiques différentes en cours de la tunnelisation. Par exemple, des applications délai-sensible ont une priorité d'être traitées plutôt que des applications non-sensibles au délai. Le trafic entrant à un point de terminaison du tunnel est classé par certain politique, et aussi mis aux différents tampons, et envoyé en utilisant différentes bandes passantes. En outre, Pour garantir la sécurité entre deux points de terminaison du tunnel, un mécanisme est aussi défini pour échanger des clés secrètes partagées. La nouveauté d'ici est de permet aux nœuds de posséder quelques paires de clés avec un nouvel concept de Key-Index. Chaque nœud a un tableau de clés (Tunneling Key Table) indexé par un nombre entier s'appelle Key-Index. Lors de l'échange des clés secrètes, un nœud choisit un ensemble des clés : (Key-Index, Public Key, Private Key). Ensuite, ce nœud envoie une partie de cet ensemble au nœud correspondant (Key-Index, Public Key). Le correspondant fait la même chose. Lors de l'encryptions des données, un nœud choisit un des clés publiques du nœud correspondant, encrypte ces données, puis les envoie vers le nœud correspondant avec l'index de cette clé. Le nœud correspondant cherche son tableau de clés et trouve la clé privée qui corresponde à l'index reçu, puis décrypte des données (Une demande de brevet basé sur ce protocole a été déposée aux États-Unis). Cette solution offre des possibilités aux administrateurs réseaux de configurer les tunnels bidirectionnels sécuritaires avant les relèves actuelles. En outre, l'utilisation de ces tunnels peut garantir une certaine qualité de service pour des sessions multimédias en cours durant une relève, car ils aident les utili-

sateurs mobiles à réserver les ressources réseaux pour leur relève imminente. Les trafics sont classés au point de terminaison du tunnel. En utilisant différentes clés secrètes partagées, un point de terminaison du tunnel encrypte des paquets et les transmet à l'autre point de terminaison du tunnel.

- Cette thèse propose également une nouvelle architecture intégrée pour les réseaux sans fil hétérogènes de la prochaine génération. La nouveauté de cette architecture est d'introduire deux nouveaux éléments de réseau : eMAP et eHAAA. Le premier combine des fonctionnalités clés de la mobilité point d'ancrage (MAP) et le domaine passerelle (WIG pour WLAN, GGSN pour l'UMTS, PDSN pour CDMA2000, DIG pour WMAN). Il est également possible pour eMAP de regrouper les fonctions du serveur LAAA [90] et Visiteur Location Register (VLR). En ce qui concerne ce dernier, eHAAA agrége les principales fonctionnalités du serveur HAAA et une base de données centrale (CD) qui contient des informations détaillées sur chaque région, par exemple l'identificateur de la région, l'adresse IP de chaque eMAP, et ses environs, l'identificateur du noeud frontière, etc. Cet eHAAA pourrait contenir également la fonctionnalité de Home Location Register (HLR), Home Subscriber Server (HSS), ou le Profil d'Utilisateur du Serveur de Fonction (UPSF). Il est à noter que le HSS ou UPSF est une base de données qui contient les profils des utilisateurs mobiles, et qui soutient les entités réseaux dans un sous-système multimedia IP (IMS) à traiter effectivement les appels. La fonction d'un HSS ou UPSF est similaire au HLR et le centre d'authentification (AuC) dans le système de GSM.
- De plus, à partir de l'architecture élaborée, cette thèse propose un mécanisme qui soutient la relève sans coupure dans les réseaux sans fil basés sur le protocole MIPv6 (Une demande de brevet a également été déposée aux États-Unis). Cette solution offre aux opérateurs de réseaux mobiles une occasion de fournir des services de communication aux abonnés qui viennent d'autres réseaux. Elle permet aux nœuds mobiles d'utiliser leurs précédentes adresses IP dans un réseau visité, à condition que l'ancien réseau d'accès et le nouveau concluent un accord pour permettre aux utilisateurs mobiles d'utiliser les tunnels pré-établis au cours de la relève. En conséquence, la signalisation utilisée pour établir les tunnels bidirectionnels est éliminée au cours d'une relève. En outre, l'originalité de ces mécanismes réside dans l'ajout de nouvelles fonctionnalités de routage au routeur d'accès, ce qui augmente l'intelligence de ce dernier en terme de sa

capacité d'analyse des paquets pour éviter de boucler sur l'ancienne adresse.

- Comme les délais de la relève de la couche deux (L2) constituent la principale composante de l'ensemble des délais de la relève, cette thèse propose une nouvelle approche qui permet la relève rapide de la couche deux dans un environnement intégrant les réseaux MIPv6 et les réseaux locaux sans fil. La nouveauté de cette approche réside dans le fait qu'elle permet aux stations mobiles de scanner rapidement les canaux et de sélectionner un canal au hasard, puis de s'associer à ce canal. L'approche proposée consiste à minimiser le nombre de canaux examinés en cours de relève et, en même temps, à reduire le temps d'attente sur chaque canal sondé. Une telle approche permet aux stations mobiles de scanner le minimum de canaux possibles à leurs portées radios. Également, les stations mobiles évitent de scanner le canal auquel elles sont attachées présentement. La raison est que des points d'accès adjacents ne peuvent pas utiliser le même canal en communications à cause de l'interférence radio. En outre, cette approche dispose d'un mécanisme qui permet aux stations mobiles d'analyser leurs canaux courants. Pour ce faire, un module appellé *Channel Analyzer* est ajouté aux fonctionnalités de la station mobile. Lors d'un scan, les stations mobiles peuvent sélectionner au hasard le premier point d'accès qui leur envoie un message de *Probe Response*. Cette technique est aussi nommée *Random Selective Scanning* ; elle permet aux stations mobiles de prendre le moins de temps possible lors d'une relève. Cela réduit le temps de relève et le nombre de pertes de paquets, ce qui correspond à l'optimisation de performance du système.

Chapitre 2. Protocole de Tunnelisation de Routers D'Accès

Ce chapitre propose un nouveau protocole de tunnelisation entre deux routeurs d'accès. Ce protocole est utilisé pour établir les tunnels bidirectionnels sécurisés (BSTs) avec un ensemble de caractéristiques minimales entre deux routeurs d'accès adjacents. Il offre également à l'administrateur du réseau l'occasion de configurer les interfaces des tunnels avant la relève effective. En outre, l'utilisation de ces tunnels peut garantir une certaine qualité de service pour des sessions multimédias en cours de relève, car les tunnels peuvent aider les utilisateurs mobiles à réserver les ressources du réseau pour leur relève imminente. Le trafic est classé à un point de terminaison

du tunnel. En utilisant différentes clés secrètes partagées, un point de terminaison du tunnel encrypte des paquets et les transmet à l'autre point de terminaison du tunnel.

Chapitre 3. Architecture Intégrée Proposée

Les progrès dans les technologies sans fil permettent aux noeuds mobiles (MNs) de bénéficier de différents réseaux sans fil tels que les réseaux locaux sans fil (WLAN), les réseaux métropolitains sans fil (WMANs), les réseaux étendus sans fil comme les réseaux cellulaires de la troisième génération (3G), etc. Ces réseaux devraient pouvoir s'intégrer les uns aux autres et fournir en tout temps un haut débit des services aux utilisateurs mobiles [3]. Ce chapitre présente une nouvelle architecture qui soutient la relève rapide et sans coupure (AFS) dans les réseaux sans fil hétérogènes de la prochaine génération. AFS étend l'infrastructure existante en intégrant les systèmes sans fil disparates qui comprennent les réseaux locaux sans fil, les réseaux métropolitains sans fil et les systèmes de la troisième génération (3G) tels que Universal Mobile Telecommunication System (UMTS) et cdma2000.

Chapitre 4. Régimes de la Relève sans Coupure Proposée

Ce chapitre fournit d'abord une revue de la littérature dans le domaine de gestion de la mobilité. Nous analysons principalement les protocoles de gestion de mobilité proposés au sein des activités de recherche de l'Internet Engineering Task Force (IETF). Ensuite, les défis pour la gestion de la mobilité dans les réseaux sans fil hétérogènes de la prochaine génération sont décrits afin de fournir une vision globale de cette recherche. Nous analysons les protocoles tels que Mobile IPv6 (MIPv6) [7], HMIPv6 [8] [9], FMIPv6 [10] [11], F-HMIPv6 [16] - [19] et le protocole de Proxy Mobile IPv6 (PMIPv6) [20] en détail. Les tendances futures dans la conception des systèmes intelligents de gestion de la mobilité sont ensuite présentées. Enfin, nous proposons des régimes qui soutiennent la relève sans coupure dans les réseaux sans fil de la prochaine génération basés sur l'architecture élaborée.

4.1 Régime proposé pour la relève sans coupure

L'idée de base est de configurer des tunnels bidirectionnels et securitaires (BSTs) entre des routeurs d'accès adjacents avant la relève effective. Les paramètres de la qualité de service et les aspects de sécurité sont précisés dans un contexte de contrat (SLA) pour chaque tunnel. Ces tunnels permettent aux opérateurs réseaux de fournir des services aux utilisateurs mobiles d'autres opérateurs, à condition que les deux parties impliquées aient signé un accord à partir duquel les utilisateurs mobiles ont la possibilité d'exploiter des tunnels pré-établis pour leurs sessions multimédias en cours de relève. Avec l'aide de tunnels bidirectionnels pré-établis, les utilisateurs mobiles peuvent utiliser leur ancienne adresse IP dans un nouveau réseau visité ou un domaine visité. Par conséquent, ceci réduit les interruptions de service au minimum durant la relève. De nouvelles fonctionalités de routage sont ajoutées au routeur d'accès. Ceci permet l'acheminement de paquets vers les utilisateurs mobiles qui ont une adresse invalide topologiquement dans le réseau attaché.

Chapitre 5. Régimes de la Relève Rapide Proposée

Basés sur le standard IEEE 802.11, les réseaux locaux sans fil (WLAN) ont connu une croissance rapide depuis quelques années. De plus en plus omniprésents, ils sont déployés comme points d'accès sans fil dans les aéroports, les campus, les centres commerciaux, etc. afin de faciliter l'accès à Internet pour les utilisateurs mobiles. Pendant ce temps, les fournisseurs de services Internet (ISPs) peuvent disposer d'une augmentation de la productivité en permettant aux utilisateurs d'itinérance d'accéder aux données dans les systèmes sans fil. Ces faits font de la gestion de la relève un problème crucial dans les réseaux locaux sans fil. Toutefois, les normes traditionnelles de l'IEEE ne fournissent pas assez de soutien nécessaire pour la relève rapide quand les noeuds mobiles se déplacent d'un point d'accès (AP) à l'autre. En conséquence, un certain nombre de régimes qui soutiennent la relève rapide ont été proposés dans la littérature. Dans ce chapitre, ces régimes sont examinés et leurs forces et faiblesses sont exposées. Ensuite, d'importantes considérations de conception pour la gestion de relai dans les réseaux locaux sans fil sont identifiés. Les problèmes potentiels de la recherche sont mis en évidence par rapport à l'amélioration de la performance et le soutien du transfert des applications en temps réel. Après avoir discuté des problèmes liés à notre recherche, nous proposons un nouveau régime qui est basé sur la couche

2 (MAC) et permet la relève rapide pour un environnement sans fil intégrant la norme de MIPv6 et la norme IEEE 802.11b. Ce nouveau régime vise à appuyer les applications en temps réel et la relève rapide lorsque des noeuds mobiles changent leur points d'attache réseautiques. Il consiste à minimiser le nombre total de canaux examinés durant une relève, ainsi que le temps d'attente pour chaque canal examiné. La performance est évaluée par des simulations dont les résultats montrent que notre proposition offre de meilleures performances, par rapport à la norme IEEE 802.11b, la norme IEEE 802.11b avec *MinChannelTime* et deux autres solutions dans la littérature : Selective scanning et AP cache, et des solutions de Neighbor Graphs.

Chapitre 6. Conclusion

Cette thèse propose d'abord un nouveau protocole de la tunnelisation entre des routeurs d'accès, qui permet aux fournisseurs de services de soutenir la relève sans coupure pour les utilsateurs mobiles venant d'autres fournisseurs de services. De plus, une nouvelle architecture intégrée est conçue pour les réseaux sans fil hétérogènes de la prochaine génération. Basés sur l'architecture élaborée, des régimes pour la gestion de la relève sans coupure sont proposés pour faciliter le soutien de la mobilité et pour fournir une qualité des services à partir de la couche IP (couche trois). Pour évaluer la performance de ces régimes, nous employons des modèles analytiques en étudiant l'impact de divers paramètres du système sans fil sur le processus de relève. Des simulations sont réalisées avec OPNET Modeler v.12.0 pour étudier les performances de la relève. Comme le temps de la relève de la couche deux est un composant important parmi les délais totaux de relève, nous avons également proposé une nouvelle approche pour garantir la relève rapide à la couche MAC dans un environement intégrant de la norme MIPv6 et IEEE 802.11 WLANs. Et l'évaluation de la performance se fait par les simulations avec le simulateur SimulX.

Plusieurs thèmes sont ouverts à la recherche future dans le domaine de la gestion de mobilité. Étant donné que les nouvelles améliorations pour MIPv6 sont portées au sein de l'IETF par des groupes de travail tous les jours à chaque instant, nous pensons que les nouveaux protocoles de relève sans coupure sont nécessaires pour non seulement réduire la latence de relève, mais aussi pour minimiser les pertes de paquets causés par le processus de transfert.

Même si les régimes SMIPv6 fournissent de meilleures performances que MIPv6 et ses extensions telles que HMIPv6, FMIPv6, F-HMIPv6, nous constatons que ces régimes sont toujours centrés hôtes. En d'autres termes, ils obligent les nœuds mobiles à signaler la gestion de la mobilité à leurs agents mères et à tous les noeuds correspondants actifs. Toutefois, dans le cas où un nœud mobile n'a pas la capacité de transmettre la signalisation de la mobilité, les protocoles de gestion de la mobilité orientés hôtes ne seront plus fonctionnels. C'est une des limites de ce travail de recherche. Comme les régimes de SMIPv6 présentent des noeuds capables de transmettre la signalisation, l'interfonctionnement avec le protocole de Proxy Mobile IPv6 (PMIPv6) doit être pris en considération dans un proche avenir.

De plus, comme la performance est évaluée à l'aide de modèles analytiques et de simulations, l'utilité de SMIPv6 ne peut être pleinement mise en œuvre. Les paramètres des systèmes sans fil ont par ailleurs en grande partie un impact sur les performances du système. Déterminer comment sélectionner les valeurs de paramètres qui permettent au protocole SMIPv6 d'atteindre son objectif de garantir zéro perte de paquets est une question difficile.

Cette thèse se concentre sur la gestion de la mobilité dans les réseaux sans fil de la prochaine génération. Toutefois, en réalité, à chaque fois qu'un utilisateur mobile tente d'obtenir des services, il doit nécessairement se soumettre à un processus d'authentification. Il en résulte plus de retards en cours de relève. En conséquence, de nouveaux mécanismes d'authentification rapides sont nécessaires ainsi que la gestion de la mobilité sans coupure.

La plupart des projets visant à concevoir des nouvelles architectures utilisent des modèles analytiques pour évaluer la performance. Cependant, comme les modèles analytiques sont toujours fondés sur un certain nombre d'hypothèses, les résultats numériques obtenus sont sujets à caution. Par conséquent, la mise en œuvre de l'installation de bancs d'essai réel sera préférable dans un proche avenir.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

3G	Third Generation
3GPP	Third Generation Partnership Project
3GPP2	Third Generation Partnership Project 2
4G	Fourth Generation
AAA	Authentication, Authorization and Accounting
AD	Administrative Domain
AFS	Architecture for Fast and Seamless roaming
AN	Access Network
AP	Access Point
AR	Access Router
ARTP	Access Router Tunneling Protocol
AS	Authentication Server
AuC	Authentication Center
AuD	Authentication Database
BAck	Binding Acknowledgment
BCE	Binding Cache Entry
BS	Base Station
BSS	Basic Service Set
BSSID	Basic Service Set Identifier
BST	Bidirectional Secure Tunnel
BTS	Base Transceiver Station
BU	Binding Update
CDMA	Code Division Multiple Access
CMR	Call-to-Mobility Ratio
CN	Correspondent Node
CoA	Care-of Address
CSMA/CA	Carrier Sense Multiple Access with Collision Avoidance
DAD	Duplicate Address Detection
DHCP	Dynamic Host Configuration Protocol
DIG	Domain Interworking Gateway
DoS	Denial of Service

EAP	Extensible Authentication Protocol
EAPOL	Extensible Authentication Protocol Over LANs
eHAAA	enhanced Authentication, Authorization and Accounting server
eMAP	enhanced Mobility Anchor Point
ESS	Extended Service Set
FBAck	Fast Binding Acknowledgement
FBU	Fast Binding Update
F-HMIPv6	Fast handover for Hierarchical Mobile IPv6
FHR	Frequent Handoff Region
FMIPv6	Fast handovers for Mobile IPv6
FNA	Fast Neighbor Advertisement
GGSN	Gateway GPRS Support Node
GPRS	General Packet Radio Service
GSM	Global System for Mobile communications
HA	Home Agent
HAck	Handover Acknowledge
HI	Handover Initiate
HIN	Home Intelligent Node
HLR	Home Location Register
HMIPv6	Hierarchical Mobile IPv6 (RFC4140)
HoA	Home Address
HSS	Home Subscriber Server
IAPP	Inter-Access Point Protocol
IETF	Internet Engineering Task Force
IMS	IP Multimedia Subsystem
IP	Internet Protocol
IPv4	Internet Protocol version 4
IPv6	Internet Protocol version 6
ISP	Internet Service Provider
L2	Layer two
L3	Layer three
LAN	Local Area Network
LBACK	Local Binding Acknowledgment
LBU	Local Binding Update

LCoA	Local Care-of Address or on-Link Care-of Address
LMA	Local Mobility Anchor
LTE	Long Term Evolution
MAC	Medium Access Control
MAG	Mobile Access Gateway
MAHO	Mobile Assisted Handoff
MAP	Mobility Anchor Point
MCHO	Mobile Controlled Handoff
MD5	Message-Digest algorithm 5
MIH	Media Independent Handover
MIPv4	Mobile IPv4
MIPv6	Mobile IPv6 (RFC3775)
MITM	Man-In-The-Middle
MM	Mobility Management
MN	Mobile Node
NA	Neighbor Advertisement
NAACK	Neighbor Advertisement Acknowledge
NAHO	Network Assisted HandOff
NAPT	Network Address and Port Translation
NAR	New Access Router
NAV	Network Allocation Vector
NB	Node B
NCHO	Network Controlled HandOff
neMAP	new enhanced Mobility Anchor Point
NG	Neighbor Graph
NGC	Neighbor Graph Caching
NGWN	Next Generation Wireless Network
N-MAG	New MAG
NS	Neighbor Solicitation
P2P	Point-to-Point
PAR	Previous Access Router
PBAck	Proxy Binding Acknowledgment
PBU	Proxy Binding Update
PCF	Packet Control Function

PCoA	Previous Care-of Address
PDA	Personal Digital Assistant
PDSN	Packet Data Serving Node
PHY	Physical layer
PKI	Public-Key Infrastructure
P-MAG	Previous MAG
PMIPv6	Proxy Mobile IPv6
PNC	Proactive Neighbor Caching
PrRtAdv	Proxy Router Advertisement
P2P	Point-to-Point
QoS	Quality of Service
RA	Router Advertisement
RAT	Radio Access Technology
RADIUS	Remote Authentication Dial-In User Service
RAN	Radio Access Network
RCoA	Regional Care-of Address
REN	Regional Edge Node
RFC	Request For Comments
RIN	Regional Intelligent Node
RO	Route Optimization
RR	Return Routability
RS	Router Solicitation
RtSolPr	Router Solicitation for Proxy Advertisement
SA	Security Association
SAE	System Architecture Evolution
SHA	Secure Hash Algorithm
SLA	Service Level Agreement
SMIPv6	Seamless handoff schemes for MIPv6
SMR	Session-to-Mobility Ratio
SNC	Selective Neighbor Caching
SSID	Service Set Identifier
TB	Tunnel Broker
TCP	Transmission Control Protocol
UDP	User Datagram Protocol

UMA	Universal Mobile Access
UMTS	Universal Mobile Telecommunications System
UNA	Unsolicited Neighbor Advertisement
VAS	Value Added Service
VLR	Visitor Location Register
VoIP	Voice over IP
UPSF	User Profile Server Function
WEP	Wired Equivalent Privacy
WIG	Wireless Interworking Gateway
WiMax	Worldwide interoperability for Microwave access
WLAN	Wireless Local Area Network
WMAN	Wireless Metropolitan Area Network
WPAN	Wireless Personal Area Network
WWAN	Wireless Wide Area Network

CHAPTER 1

INTRODUCTION

With the rapid progress made in wireless technology, the number of mobile users has been incredibly growing in recent years. According to the statistics of 3G Americas, the number of worldwide cellular subscriptions has increased to 3.31 billion by the end of December 2007, among which the subscriptions using the technologies of *global system for mobile communications* (GSM) and *universal mobile telecommunications system* (UMTS) have reached to 2.88 billion while the subscriptions using *code division multiple access* (CDMA) technologies have risen to 376 millions [1]. As more and more wireless users account for a majority of the Internet population today, mobility management (MM) becomes an important research challenge in the wireless system design.

On one hand, advanced wireless technologies enable mobile nodes (MNs) to be equipped with multi-mode radio interfaces, thus wireless users have opportunities to benefit from disparate mobile communication systems, such as wireless personal area networks (WPANs), wireless local area networks (WLANs), wireless metropolitan area networks (WMANs), wireless wide area networks (WWANs) like third generation (3G) cellular systems, etc. These networks tend to complement each other, and integrate one another to provide ubiquitous and high data-rate services to roaming users [2], [3].

On the other hand, a number of mobility management protocols have been designed by standardization activities within the Internet Engineering Task Force (IETF) [4], such as mobile IPv4 (MIPv4) [5], [6], mobile IPv6 (MIPv6) [7], hierarchical mobile IPv6 (HMIPv6) [8], [9], fast handovers for mobile IPv6 (FMIPv6) [10]-[15], fast handover for hierarchical mobile IPv6 (F-HMIPv6) [16]-[19], proxy mobile IPv6 (PMIPv6) [20]-[26], localized proxy mobile IPv6 [27] and fast handovers for proxy mobile IPv6 (F-PMIPv6) [28], [29] and fast localized proxy mobile IPv6 (FLPMIPv6) [30], etc. Consequently, wireless networks are ready to provide mobility support for roaming users across the same or different access technologies.

However, as the demands from wireless users for high-speed Internet access and

multimedia applications increase, it is necessary for next generation wireless networks (NGWNs) to provide mobile user with the support of seamless mobility and quality of service (QoS) provisioning during handoff. This adds more complexity to the mobility protocol design, since network connectivity, session continuity and quality of the session need to be taken into account simultaneously while handover mobile users' ongoing sessions from one network to another. Note that these involved networks can deploy either homogeneous or heterogeneous radio access technology.

In the following paragraphs, we introduce some basic concepts and definitions in the domain of mobility management. Such concepts and definitions will assist us to better understand the research background of seamless mobility management problem. Followed by the definitions, we describe the motivations and challenges of this research. And then we define our research objectives, and describe the contributions and originalities of this research. Finally, we briefly outline the thesis plan.

1.1 Basic concepts and definitions

Mobility management (MM) enables wireless communication systems to locate mobile nodes (MNs) for call or data delivery and to maintain connections as the MNs change their network attachment points (or access points (APs)) [31], [32]. It typically consists of two components: *location management* and *handoff management.*

Location management enables the wireless network to discover an MN's current AP for call or data delivery. And it can be further divided into two stages: *location registration* (or *location update*) and *data delivery* (or *call delivery*). The former requires mobile users to periodically notify the network of their current location, and allows the network to authenticate the users and revise the users' location profile. While the latter queries the network for the user's location profile and find the mobile user's current position in the network [31].

The techniques used for location management include database architecture design, reducing the signalling between different network components, pointer approach [33], etc. The research challenges with respect to location management are listed as follows [31]:

- Support effectively the continuously increased population of mobile users;
- Provide a level of security and privacy that satisfy both mobile users and net-

work service providers;

- Dynamic and efficient update users' profile and location databases;
- Reduce the querying delays incurred by data delivery or paging process;
- Design efficient terminal paging methods;
- Minimize the paging delays.

Roaming implies formal agreements between network operators that allow MNs to obtain connectivity from foreign networks. And it allows mobile users to communicate their identity to the visiting access network (AN), thus activates inter-AN agreements. Afterwards, the foreign network emulates the mobile users' home networks and offers them communication services [34]. *Handoff* (or *handover*) is a process by which an active MN changes its network attachment point or when such a change is attempted. During handoff, ongoing sessions are interrupted and such interruption need to be minimized by the network [34].

Handoff can be divided into two categories: *network controlled handoff* (NCHO) and *mobile controlled handoff* (MCHO). NCHO enables the network to generate new connections for mobile terminals, find new resources for the handoff and perform any additional routing operations. In other words, handoff decision is made by a network element [35]. In MCHO, the handoff is decided and controlled by mobile terminals themselves.

In addition, a handoff decision usually involves some measurements about when and where to handover to. In this case, handoff can be classified into *mobile assisted handoff* (MAHO), *network assisted handoff* (NAHO) and *unassisted handoff* [34]. During the MAHO, handoff related information are collected and measured by MNs, and used by ARs to make handoff decision. That is, MAHO enables MNs to find new network resources and allows the network to approve the handoff. Generally, the MAHO involves feedback from MNs as part of the handoff process. The feedback includes signal level from neighbor cells and downlink signal quality, etc. The NAHO allows the network to help MNs to make handoff decision while unassisted handoff implies no assistance from either the network or mobile terminals.

Handoff can also be classified into *intra-* and *inter-technology* handoff. The former indicates a handover between equipment of the same technology while the latter means a handover between equipment of different technologies [34].

Handoff can be grouped into *horizontal* and *vertical handoff* as well [34]. The former implies that MNs move between the APs of the same type in terms of radio coverage, data rate and mobility protocol. For example, handoff between the systems of UMTS or between WLANs is horizontal handoff. And vertical handoff takes place when the MNs move between the APs of different type. An example of such handoff is that an MN moves its association from a UMTS to a WLAN [34].

Handoff management enables the network to maintain mobile users' network connectivity as they change the network APs. It can be further divided into *handoff initiation* (or *handoff detection*), *new access network discovery* and *data path modification* stages. During the first stage, the needs of handoff are detected and signalized by the mobile terminal, a network agent, or the network itself. During the second stage, the mobile terminal discovers a new access network and obtains network connectivity. During the last stage, the network needs to modify and maintain the data path from the old connection path to the new one according to some agreed service guarantees [31]. The research issues about handoff management are listed as follows:

- Detect efficiently and quickly the needs of handoff;
- Find the best connected network;
- Minimize signalling load/overhead on the wireless network;
- Optimize the route modification for each communication session;
- Improve the network resource utilization, especially bandwidth reassignment;
- Reduce packet loss rate and jitter during handoff;
- Guarantee the security and privacy of communications in progress;
- Minimize the interruption to sessions in progress;
- Provide QoS for mobile terminals, especially those with real-time multimedia applications.

Besides the categories of location management and handoff management, mobility management can also be classified into *macro-* and *micro-mobility management* according to the size of mobility domain. *Macro-mobility* (or *global mobility*) indicates the mobility over a large geographic area. It usually includes mobility support, new

access network discovery, new IP address configuration, verification and registration when MNs move between different administrative domains (ADs) [34]. MIPv4 [5] [6] and MIPv6 [7] are considered as the protocols for such mobility management. *Micro-mobility* implies the mobility across small geographic areas, and it usually involves the movements within an IP administrative domain [34]. Under the circumstances, signalling messages related to the movements are confined to a visiting access network or a local administrative domain. For example, FMIPv6 protocol [10] [11] limits the signalling with respect to mobility management to local access routers (ARs): previous AR (PAR) and new AR (NAR). And HMIPv6 [8], [9] and F-HMIPv6 [16]-[19] protocols restrict the mobility related signalling to a local domain, managed by a mobility anchor point (MAP).

Relating to the mobility management, some network elements need to be mentioned. Such as a *mobile node* (MN) is an IP-capable device that can change its network AP while still being reachable via its home address (HoA) [7]. A *correspondent node* (CN) is a peer node with which an MN is communicating [7]. A CN may be either mobile or stationary. An *access router* (AR) is a router dwelling on the edge of an access network and connected to one or more APs. It offers IP connectivity to MNs, acting as their default routers. And the AR may include intelligence besides the functionalities of ordinary routers [34]. An *access point* (AP) is a layer two (L2) device connected to one or more ARs and offers wireless link connection to MNs [34]. A *home agent* (HA) is a router on an MN's home link with which the MN has registered its current care-of address (CoA). When the MN moves away from its home network, the HA intercepts the packets destined to the MN's HoA, encapsulates and tunnels them to the MN's new location, which is presented by the MN's CoA [34]. A *mobility anchor point* (MAP) is a router located in a visiting network/domain and used by the MNs as their local HAs [8] [9].

According to the performance and functional aspect, handoff can be sorted into *smooth*, *fast* and *seamless handoff*. *Smooth handoff* is a handoff technique that mainly aims to minimize packet loss rate, without explicit concern for additional delays in packet forwarding [34]. *Fast handoff* aims to reduce handover latency, without explicit interest in minimizing packet losses during handoff [34]. And *seamless handoff* attempts to guarantee no change in service capability, security, or quality [34]. The latency pertaining to the handover process and packet losses during handoff are the critical factors for seamless handoff [36], because seamless handoff aims to offer a given

QoS [37] and provide mobile users with imperceptible session interruption while they change their network attachment points.

During handoff, there is always a period during which an MN cannot send or receive packets due to link switching and IP protocol operations. This period is called *handoff latency*. It is also defined as the time difference between the moment when the MN sends or receives the last packet through its associated AR (also called the PAR) and the moment when it can send or receive the first packet via the new AR (NAR). Such latency usually consists of movement detection, new CoA configuration, verification and location registration (or binding update) procedures [10] [11].

In the literature, handoff latency usually comprises the delays pertaining to layer two (L2) and layer three (L3) handoff processes. *L2 handoff* refers to a process by which an MN changes its link-layer connection while *L3 handover* commonly takes place after a L2 handoff, in which an MN detects the change of network prefix, and configures a new IP address to regain network connectivity [7].

Concerned with other QoS parameters, *throughput* is defined as the amount of data from a source to a destination processed by the protocol [34]. *Packet loss rate* is defined as the percentage of the number of received packets at a destination node over the number of packets sent by a source node. *Jitter* refers to the time interval between successive packets.

Here we also define some handoff related signalling messages, such as a *beacon*, which is a control message broadcast by an AP to inform all neighboring nodes of its presence in the network [34]. Typically, before obtaining connectivity from a visiting network, an MN receives beacons from adjacent APs. Such beacons assist the MN to discover APs within range and to make handoff decisions. In addition, with the help of candidate access router discovery (CARD) protocol [38], the MN can resolve or formulate a new on-link CoA based on the mapping between the identity of an AP (AP-ID) and the IP address of an AR. *Link-layer trigger* (or L2 trigger) is the information from the link layer to notify the network layer of the detailed handoff events at the link layer [34]. For example, FMIPv6 protocol defines several triggers such as *link up trigger*, which indicates that the MN establishes a connection with an AP; *link down trigger*, which means that the MN loses its connection with its associated AP; and *L2 handover start trigger*, which signifies that the MN starts a L2 handover to a new AP [10] [11].

1.2 Motivations and research challenges

Recently, the design of protocols to support seamless mobility and QoS provisioning becomes a hot topic in all-IP-based next-generation heterogeneous wireless networks. Since in such networks, mobile users must be provided with the capability of conveniently roaming between various operators, across different access technologies. In addition, next generation wireless networks (NGWNs) tend to support value added service (VAS), such as short message service (SMS), multimedia messaging service (MMS), general packet radio service (GPRS), IP based telephony and multimedia services, etc. However, this imposes new complexity in the wireless system design, because provisioning of such services in the NGWNs brings about new QoS requirements. For example, multimedia applications require an optimal performance of the wireless systems in terms of handoff delay, end-to-end latency, packet loss rate, and jitter, etc.

In addition, the provisioning of synchronous real-time applications, such as voice over IP (VoIP) and video-conferencing over IP, place new challenges for QoS provisioning. For example, the real-time applications such as voice or video over IP set severe temporal requirements on mobility management protocols: the handoff latency for seamless handover scenarios is limited to less than $100ms$, and jitter disturbances are restricted to less than $50ms$. Note that $100ms$ is about the duration of a spoken syllable in real-time audio traffic [39].

In order to find out new solutions for the problem of IP layer fast and seamless mobility management, it is necessary to analyze the components of handover latency. It is taken for granted that handover mechanisms in IP layer have two objectives: reducing the handover delays and avoiding packets losses incurred by the handoff process. According to these objectives, handoff protocols are categorized into *fast handoff*, *smooth handoff* and *regional registration* [41]. Figure 1.1 shows a typical handover process [41].

As shown in Figure 1.1, a handover period can be partitioned into three phases from point 1 to point 4. After the moment of starting a handoff, an MN needs to discover a new link to associate with, usually this process implies link switching (or link layer handoff) processes. To become reachable in the Internet, the MN also needs to find out a new default router after link switching. This is done through neighbor discovery procedures [42], in which the MN sends out router solicitation (RS)

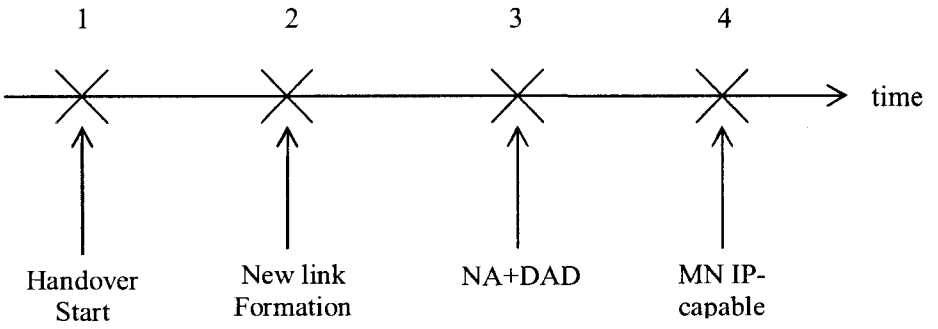

Figure 1.1 Typical Handover Process

messages and the ARs within range reply with router advertisement (RA) messages. Upon discover new AR, the MN formulates a new IP address, called care-of address (CoA) in the new visited network. And the MN then announces its presence by sending neighbor advertisement (NA) messages on the new link. Subsequently, the MN performs duplicate address detection (DAD) process [43] to verify the uniqueness of the new CoA. After successful registration with its home agent (HA), the MN becomes IP-capable in the visiting network.

Regional registration means registration done locally to the visiting domain either via a gateway foreign agent (GFA) [44] or a mobility anchor point (MAP) [8] [9]. Such registration is used to reduce the time from point 3 to point 4 during handover. *Fast handover* approach (e.g. FMIPv6 [10] [11]) tries to shorten the delay from point 2 to point 3 while *smooth handoff* aims to reduce packet drops during the period from point 1 to point 4 when an MN cannot send or receive any packets [41].

As seamless handoff attempts to minimize service disruption during handoff, it implies minimal handoff delay, low packet loss, less signalling overhead incurred by the handoff process. In other words, seamless handoff is defined as a handover procedure which does not cause any degradation of service noticeable by a mobile user [45]. Hence, the delays and packet losses from point 1 to point 4 are required to be taken into account in the design of new IP layer seamless handoff protocol. In general, the handover latency contains the following elements:

- **Link layer handoff delay** is the time taken to establish link layer connectivity.
- **Movement detection delay** is the duration for an MN to receive RAs from new ARs within range. Such delay depends on how soon the MN can detect its

movements beyond the coverage area of its default AR and initiates a handover.

- **New CoA configuration and verification delays** are the time taken for an MN to configure a new CoA on the link and verify the uniqueness of the new CoA through the DAD process. Typically, the MN can configure a new CoA either in a stateless way [43] or in a stateful way through a dynamic host configuration protocol (DHCP) server [46].

- **Registration delay** is the time difference between the moment when an MN sends a binding update (BU) message to its home agent (HA) via the previous AR (PAR) and the moment when the MN receives the first packet via the new access router (NAR).

As seen from the components of handoff latency, supporting seamless mobility with QoS provisioning is a challenging issue in the wireless network design. In addition, mobility support can be provided from various layers of TCP/IP protocol stack reference model [40], which include link layer, IP layer, transport layer, cross-layer (layer 2 + layer 3 or layer 3 + layer 4), etc.

1.3 Research objectives

The objective of this research is to develop new fast and seamless mobility management protocols for IPv6-based next-generation wireless networks. More specifically, new schemes for handoff management in mobile IPv6-based wireless networks for network layer mobility support, new schemes for handoff management in a MIPv6/WLANs environment for link layer mobility support.

In other words, the principal objective of this research is to propose new mobility management protocols that guarantee minimal service disruption during handoff for mobile users' ongoing multimedia sessions. More specifically, we design handoff management schemes to minimize handoff related signalling overhead, handoff latency, packets loss rate and jitter. For further details, this thesis takes the following aspects into account:

- Propose new protocol that enables intelligent and dynamic configuration of IP tunnelling between access routers;

- Propose new integrated architecture for IPv6-based NGWNs;

- Develop new protocol that empowers access routers to discovery their neighborhood, to exchange information about their capabilities, and to establish trust relationship between them;
- Develop novel seamless handoff mechanisms for MIPv6-based wireless and mobile networks;
- Design new method to protect sessions in progress during handoff;
- Evaluate the efficiency of the proposed handoff schemes with analytical models and simulations;
- Design link-layer fast mobility management solution in an environment that integrates the protocol of MIPv6 and IEEE 802.11-based WLANs;
- Analyze the performance of the proposed L2 handoff schemes with simulations.

1.4 Contributions and originalities

Major contributions and originalities of the thesis are listed as follows:

- This thesis proposes a new access router tunnelling protocol (ARTP), which enables establishing a tunnel with a set of minimal characteristics between two routers (or nodes) in a network. This patent-pending solution provides network administrator the opportunity to configure bidirectional secure tunnels (BSTs) prior to actual handoff. In addition, using such tunnels can guarantee certain QoS for ongoing multimedia sessions during handoff, because the pre-established tunnels can assist roaming users to reserve network resources for their impending handoff. Moreover, the incoming traffic is classified at tunnel endpoint. Using different shared secret key, one tunnel endpoint encrypts incoming packets and forwards them to the other tunnel endpoint.
- This thesis also proposes a novel integrated architecture for next-generation heterogeneous wireless networks. The novelty of this architecture is represented by the introduction of new network elements: eMAP and eHAAA. The former combines key functionalities of mobility anchor point (MAP) and domain gateway, for example, wireless interworking gateway (WIG) for WLAN, gateway GPRS support node (GGSN) for UMTS, packet data serving node (PDSN)

for cdma2000, domain interworking gateway (DIG) for WMAN, etc. It is also possible for eMAP to aggregate the functionality of local authentication, authorization and accounting (LAAA) server [90] and visitor location register (VLR). As to the latter, eHAAA aggregates key functionalities of home AAA server (HAAA) and a central database (CD) that contains detailed information about each region, e.g. region identifier, each eMAP's IP address and its neighborhood, regional edge node (REN) identifier, etc. The eHAAA may also contain the functionality of home location register (HLR), home subscriber server (HSS), or user profile server function (UPSF). Note that HSS or UPSF is a major user database to support the network entities of the IP multimedia subsystem (IMS) to handle calls, similar to the HLR and authentication center (AuC) in the system of GSM.

- Based on the designed architecture, this thesis proposes seamless handoff schemes for MIPv6-based wireless networks. This patent-pending solution provides mobile network operator an opportunity to deliver communication services to roaming subscribers from other networks. It allows an MN to utilize both its previous care-of address and new care-of address in a visiting network, on the condition that the previous access network and the new one make an agreement to allow such node to employ pre-established tunnels during handoff. As a result, the signalling overhead for establishing bidirectional tunnels is removed from the handoff process. Additionally, the novelty of seamless handoff schemes is also presented by adding new routing functionalities at access router. Thus access router in our system is provided with more intelligence.

- As layer two (L2) handoff delays are the major component of the overall handoff latency, this thesis proposes new fast MAC layer handoff mechanism for a MIPv6/WLANs environments. Such mechanism consists of minimizing the number of scanning channels during handoff, at the same time, reducing the probe-waiting time on each probed channel.

1.5 Thesis outline

This thesis is organized as follows. In Chapter 2, an access router tunnelling protocol (ARTP) is proposed, which allows access routers to dynamically configure bidirec-

tional tunnels before actual handoff. As a result, each tunnel is specified with a set of minimal QoS parameters. In Chapter 3, after a survey of integrated and interworking architectures for next-generation wireless networks in the literature, we propose a new integrated architecture for IPv6-based next generation heterogeneous wireless systems. And then, based on the designed architecture, seamless handoff schemes are proposed in Chapter 4. To evaluate the performance of this proposal, we adopt two analytical models, within which different wireless system parameters on the handoff performance are investigated. Moreover, simulations are carried out with OPNET Modeler v.12.0 to analyze the efficiency of the proposal, in terms of control traffic sent during handoff, end-to-end latency during handoff, etc. In Chapter 5, upon a comprehensive review of fast MAC layer handoff schemes, we propose a new fast MAC layer handoff management scheme for 802.11-based WLANs. And simulations are executed with the simulator SimulX to analyze the performance of this proposal in an environment integrated the standards of MIPv6 and IEEE 802.11b. Finally, the research work of this thesis is summarized in Chapter 6 and future work is also pointed out.

CHAPTER 2

ACCESS ROUTER TUNNELLING PROTOCOL (ARTP)

This chapter proposes a new access router tunnelling protocol (ARTP), which are used to establish bidirectional secure tunnels (BSTs) with a set of minimal characteristics between two adjacent access routers. Such protocol provides network administrator an opportunity to configure the interfaces of tunnels prior to actual handoff. In addition, the utilization of such tunnels can guarantee certain quality of service for multimedia sessions during handoff, as the pre-established tunnels can assist roaming users to reserve network resources for their impending handoffs. Moreover, incoming traffic are classified at each tunnel endpoint. Using different shared secret key, the tunnel endpoint encrypts the incoming packets and forwards them to the other tunnel endpoint.

2.1 Basic concepts and definitions

In order to well understand the concept of tunnelling, we introduce some basic concepts and definitions to clarify the background of tunnelling techniques.

A *tunnel* is a forwarding path between two nodes on which the payloads of packets are packets that undergo encapsulation [47]. *Tunnel header* is the header pre-pended to the original packet during encapsulation. It specifies the tunnel end-points (including the entry-point and the exit-point) as source and destination. *IPv6 tunnelling* is a technique to establish a virtual point-to-point (P2P) link between two IPv6-capable nodes for transmitting data packets as payloads of IPv6 packets. An IPv6 tunnel is usually unidirectional in which the tunnelled packets flow in one direction from the tunnel entry-point to the exit-point. Figure 2.1 shows an example of unidirectional tunnel from a previous access router (PAR) to a new access router (NAR).

Acting as a tunnel entry-point, one node encapsulates the original packets received from other nodes (or from itself) and forwards the tunnelled packets through the tunnel while the other node, tunnel exit-point, decapsulates the received tunnelled packet and forwards the resulting original packets towards the destination node. An

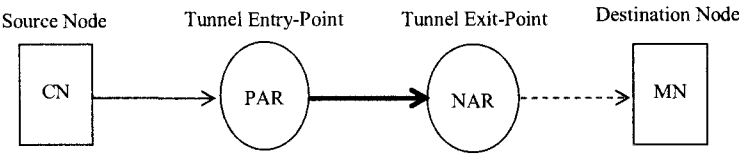

Figure 2.1 A Unidirectional Tunnel from a PAR to a NAR

access router can be either a tunnel entry-point or a tunnel exit-point. In addition, bidirectional tunnels usually consist of a forwarding tunnel and a reverse tunnel. This can be achieved by configuring two tunnels, each in opposite direction to the other. In other words, the entry-point of one tunnel is configured to be the exit-point of the other tunnel [47]. Figure 2.2 shows an example of bidirectional tunnels between a PAR and a NAR.

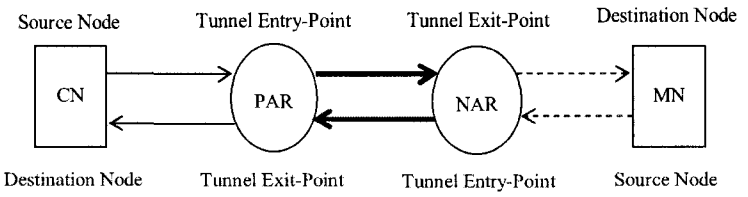

Figure 2.2 Bidirectional Tunnels between a PAR and a NAR

2.2 Requirements of access router

The access router (AR) in the proposed ARTP protocol should have features as follows:

- Support reverse tunnelling technique, since outgoing traffic from an MN to a CN goes through a reverse tunnel from a NAR to a PAR before the CN updates its binding cache entry (BCE) for the MN.
- Be programmable, because some codes with respect to new routing policies will be written and added to the functionalities of an AR that involves in the handoff process.

- Support fragmentation of packets according to the specific tunnel maximum transmission unit (MTU), which is the path MTU minus the size of the tunnel header.
- Support IPv6 encapsulation technique [47].
- Depending on the security policies exploited by the AR, reverse tunnelled packets may be discarded unless accompanied by a valid encapsulating security payload (ESP) header [48]. Moreover, to protect the network and CNs from malicious nodes masquerading as an MN, the AR must support the authenticated reverse tunnelling.
- Must verify that the source address in the tunnel IP header is the MN's previous care-of address (PCoA) when a PAR decapsulates the tunnelled packets from a NAR, this avoids that any node in the Internet sending traffic through ARs, but escape ingress filtering limitations. This simple check forces the attacker to know the current location of the real MN and be able to defeat ingress filtering. If ESP is used to protect the reverse-tunnelled packet in tunnel mode, this check is not necessary.
- Support insertion and extraction of session tokens from tunnelled packets.
- Must have enough buffer space for packets on-the-fly and somehow intelligence to analyze the received packets.

2.3 Proposed access router tunnelling protocol

In this thesis, we propose a protocol called access routers tunnelling protocol (ARTP), which aims to configure bidirectional secure tunnels (BSTs) between access routers (ARs) before actual handoff. This new signalling protocol is used to configure tunnel parameters between two end-points. In this protocol, the Internet control message protocol (ICMPv6) type messages are defined and used to transport the signalling of ARTP among ARs.

The purpose of this protocol is to negotiate a set of parameters for each preconfigured tunnel between two tunnel end-points. These parameters include the following aspects:

- QoS parameters such as guaranteed bandwidth, packet loss rate, etc.
- Security policies like authentication method, encryption method, etc.
- Traffic classification and the mapping strategy between traffic classes of different protocol or layer.
- Buffering mechanism to minimize the number of lost packets.

Since we define ICMPv6 type messages such as *Tunnel Request*, *Reverse Tunnel Request*, *Tunnel Reply*, *Tunnel ACK*, *Tunnel NACK* and *Tunnel Bye* to transport the signaling between tunnel end-points, the above mentioned parameters are encapsulated into the ICMPv6 type messages as independent sub-options.

Within the protocol ARTP, tiny client codes, called tunnel broker (TB) are developed to have charge of all communications between involved ARs. In order to setup a unidirectional tunnel, the requesting TB sends a *Tunnel Request* message to its peer. Such request proposes a set of needed characteristics for the desired tunnel. Upon receiving the *Tunnel Request*, the requested TB negotiates the conditions with the requesting TB by sending *Tunnel Reply* messages. The functionalities of the brokers is programmed and deployed at each AR. Once the requesting AR sends to its peer a *Tunnel Acknowledgement (ACK)* message indicating its acceptance or refusal of the tunnel conditions, the negotiation is complete. In the following sections, we will elaborate the proposed protocol in details.

2.3.1 Tunnel setup procedure

Once an AR acting as a tunnel entry-point requires a tunnel for a particular service, its TB initiates the tunnel setup procedure. First, this AR (called the PAR in our example) builds a *neighbor table* by executing the *neighbor discovery* procedure, in which each IPv6-capable node on the same link can learn about other nodes' presence, determine their link layer addresses, find routers within range, and maintain reachability information about the paths to active neighbors [42]. Such neighbor table provides us a local vision about potential ARs with which the current AR can establish tunnel(s). Table 2.1 shows an example of neighbor table.

Note that the Flag *U* indicates that the interface to be used is up and the IP address is available. On the contrary, the flag using *D* means that the interface of the AR is down.

Table 2.1 an Example of Neighbor Table

AR-ID	MAC Addr	IP Addr	AP-ID	Net-Prefix	Metrics	Flags	Ref.	Use
AR1	5A:42:C1	2001:106:1100::1	AP1	2001:106:1100	1	U	0	1
AR1	5A:42:C1	2001:106:1100::1	AP2	2001:106:1100	1	U	0	1
AR2	90:B3:42	2001:106:2300::1	AP3	2001:106:2300	1	U	0	1
AR2	90:B3:42	2001:106:2300::1	AP4	2001:106:2300	1	U	0	1
AR3	04:69:AA	2001:106:2700::2	AP5	2001:106:2700	1	U	0	1
AR3	04:69:AA	2001:106:2700::2	AP6	2001:106:2700	1	U	0	1

Once the requesting TB successfully creates its *neighbor table*, it selects one entry (or one AR) from this table, and sends a *Tunnel Request* message to the selected AR. Generally, the proposed protocol of ARTP consists of three phases as follows:

- *Command* phase in which the requesting tunnel broker (TB) sends a *Tunnel Request* message in order to establish or update a tunnel.
- *Negotiation* phase in which the two involved TBs negotiate the conditions of the tunnel, like security policies, traffic classes, QoS parameters, context transfer security blocks, type of payment, network resource reservation, buffering mechanism, etc. The negotiation is carried out by exchanging the *Tunnel Reply* messages between the involved TBs.
- *Response* phase in which one TB sends to its peer a *Tunnel Acknowledgement (ACK)* indicating its acceptance or refusal of the conditions of the tunnel.

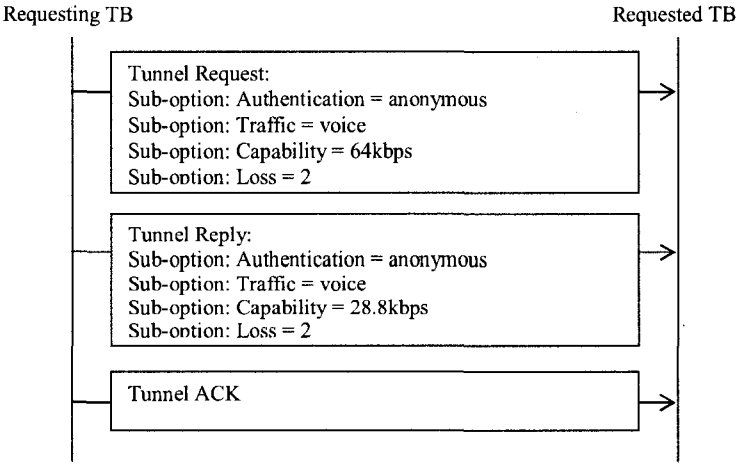

Figure 2.3 An Example of Symmetric Tunnel Setup Process

The proposed ARTP defines two kinds of tunnels: *symmetric* and *asymmetric* tunnels. The former implies that the conditions of the forwarding tunnel and the reverse tunnel have the same parameters. As to the latter, it indicates that the conditions for the forwarding tunnel and the reverse tunnel have different parameters. Figure 2.3 illustrates an example of symmetric tunnel setup process.

The *command* phase begins when the requesting TB sends a *Tunnel Request* to the requested peer. In Figure 2.3, the requesting TB asks for creating a tunnel for the voice traffic, the authentication type is anonymous, the maximum lost packets is 2 and the desired bandwidth is $64kbps$. The requested TB responds with its own available parameters with the guaranteed bandwidth is $28.8kbps$, wrapping in the *Tunnel Reply* message. Finally, the requesting AR accepts the proposal and sends a *Tunnel Acknowledgement (ACK)* to its peer. Upon receipt of *Tunnel ACK* message with acceptance, the requesting TB adds one entry into its *Forwarding Tunnel Table*, shown in Table 2.2.

Table 2.2 an Example of Forwarding Tunnel Table

No.	Entry MAC	Entry IP	Exit MAC	Exit IP	Auth	Life Time	BW	Loss	Traffic
1	5A:42:C1	1100::1	B3:42	2300::2	A	30s	28.8	2	voice

Note that the lifetime of $30s$ is the default lifetime to maintain the validity of the existing tunnels. Such value can be determined by network administrator, or depend on the interval of the neighbor discovery process. In case of establishing symmetric tunnels, the *Reverse Tunnel Table* is also build and shown in Table 2.3.

Table 2.3 an Example of Reverse Tunnel Table

No.	Entry MAC	Entry IP	Exit MAC	Exit IP	Auth	Life Time	BW	Loss	Traffic
1	B3:42	2300::2	5A:42:C1	1100::1	A	30s	28.8	2	voice

In case of asymmetric tunnels, the requesting TB must send a *Reverse Tunnel Request* to its peer after establishing the forwarding tunnel. Such message requests the requested TB to initiate the reverse tunnel setup procedure. Then the requested TB acting as the tunnel entry-point starts the aforementioned tunnel setup procedure. As a result, the requesting TB adds one entry in its *Reverse Tunnel Table*, and the requested TB adds an entry in its *Forwarding Tunnel Table*, respectively.

After the requesting TB sends a *Tunnel Request* to the requested TB, the two involved TBs negotiate the conditions of the tunnel during the *negotiation* phase.

The potential aspects to be negotiated are QoS related parameters, security policies, network resource reservation, buffering mechanism, etc. Figure 2.4 shows the bidirectional tunnel setup process.

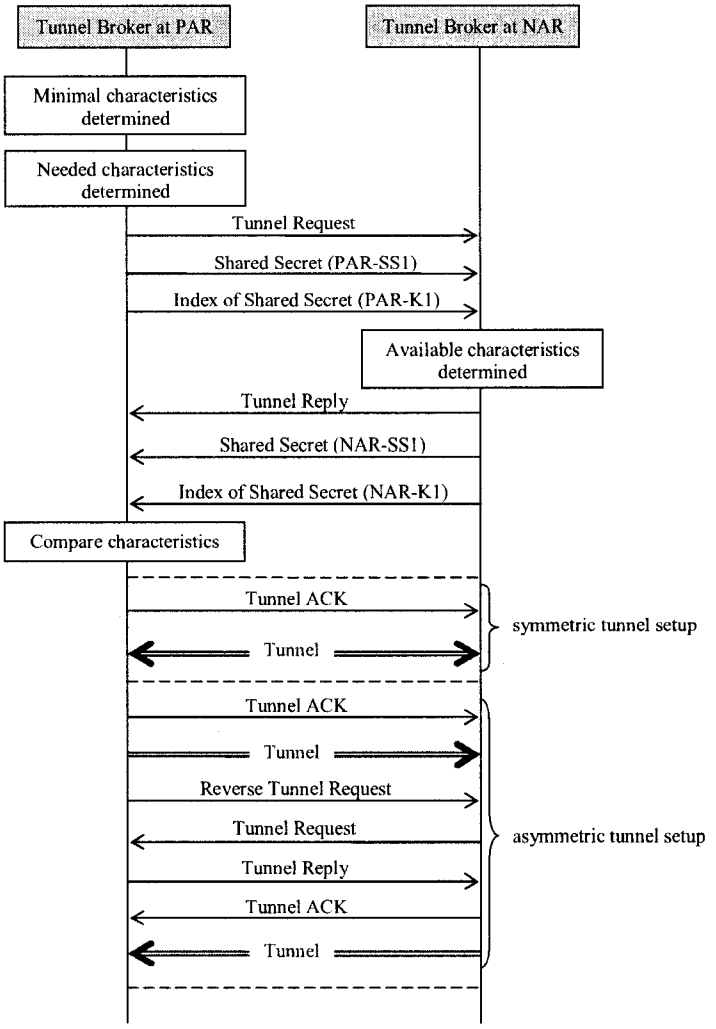

Figure 2.4 Bidirectional Secure Tunnels Setup Process

The node (AR in our case) comprises a tunnelling protocol module (called TB in this thesis) that determines a first set of desired characteristics and comprising a sub-option indicating a need for authentication characteristics. The TB at one tunnel end-point (PAR in our example) sends a *Tunnel Request* message comprising the first set of characteristics and sends a shared secret key with an index value thereof. The other tunnel end-point (NAR in our example) receives this request and

determines available characteristics, and then its TB sends a *Tunnel Reply* with the second set of desired characteristics to the requesting TB, along with a shared secret key and a corresponding key index thereof. The requesting TB compares the received set of parameters and verifies if the second set of characteristics is at least equal to the set of minimal characteristics. If so, the requesting TB sends a *Tunnel ACK* message; otherwise, it sends a *Tunnel Negative Acknowledgement (NACK)* message. The shared secret key is combined together and used to encrypt data, and the index value indicates which shared secret is used to encrypt the data.

2.3.2 QoS issues

Since quality of service allows network administrators to use their existing resources efficiently and to guarantee those critical applications receive high-quality service, without having to expand as quickly, or even over-provision, their networks, it is important to take those QoS related parameters into account during negotiation. In addition, it is undeniable that with these contracted parameters; network administrators may have more opportunities to better control over their networks, to reduce costs, and to improve customer satisfaction.

In addition, different applications have different requirements with respect to the way of handling their traffic in the network. Applications generate traffic at varying rates and generally require the network to be able to carry traffic at the rate at which they generate it. And certain applications are more or less tolerant of traffic delays in the network and of variation in traffic delay (jitter) and certain applications can tolerate some degree of traffic loss while others cannot. These QoS requirements may be expressed as follows:

- Bandwidth (BW): the rate, at which an application's traffic must be carried by the network,
- Traffic delay: the latency that an application can tolerate in delivering the data packet,
- Jitter: the delay variation in successive arrived packets.
- Loss: the percentage of lost packets, etc.

2.3.3 Security issues

In order to protect the information of IP layer and upper layer, a set of security aspects need to be considered such as access control, connection integrity, data origin authentication (or data integrity), protection against replay attack, confidentiality (encryption), and limited traffic flow confidentiality, etc. The proposed ARTP focuses on a set of algorithms used for authentication and exchange of tunnelling keys among ARs. In other words, during negotiation, the two involved ARs discuss authentication method, and algorithms used for creating tunnel keys like what is defined in [50]. The security policies include the key management and traffic protection. The selection of methods used to generate the tunnelling key, authentication method for protecting the headers and encryption method for encapsulating secure payload are negotiated while configuring the tunnel parameters. In addition, different algorithms could be used according to different traffic class to protect the outgoing packets at the tunnel entry-point.

2.3.4 Traffic classification

Network service providers could offer various services which are classified into different traffic type with corresponding guaranteed QoS, thus traffic classification becomes important in IP networks. The advantage of traffic classification is that many traffic flows can be aggregated to a small number of classes, thus simplifies the processing and storage associated with packets classification and conditioning.

Furthermore, there is no signalling state or related processing required in the differentiation, since QoS is invoked on a packet-by-packet basis. Moreover, in the IPv6 header, there is an 8-bit *Traffic Class* field available for use by originating nodes and forwarding routers to identify and distinguish between different classes or priorities of IPv6 packets. Such field may be used for the assignment of Precedence, Delay, Throughput and Reliability. Another way is to use the 20-bit *Flow Label* field to differentiate the data flow. Note that the agreement must be made on what sorts of traffic classifications are most useful for IP packets before the negotiation phase. In addition, ARs must be equipped with intelligence to decide which packet goes into which class and forward these packets according to their priority.

2.3.5 Resource reservation

Resource reservation enables Internet applications to obtain different QoS. It is well-known that different applications impose different network performance requirements. Some applications, including the more traditional interactive and batch applications, require reliable data delivery but do not impose any stringent requirements on the delivery delay. However, applications such as video conferencing, IP telephony, or other forms of multimedia communications are delay-sensitive, but unnecessary to guarantee their reliability. The proposed ARTP requests the tunnel end-point to reserve certain network resource for communications in order to guarantee seamless mobility, especially for mobile users' handoff with multimedia sessions in progress.

2.3.6 Buffering mechanism

To minimize packet losses, compromise is made between the handover latency and the number of lost packets. Hence, during negotiation, the two tunnel end-points negotiate the parameter of buffer size in terms of the percentage of its buffer space. At the end of negotiation, the two ARs make agreement with each other, and configure their tunnels based on the agreed parameters. At any time, an AR may re-start an ARTP signalling session to modify the conditions of a tunnel in case of timeout or one of its partners becomes unreachable.

2.4 Tunnelling process

When tunnelling technique is required, the tunnel entry-point encapsulates the original packets, and tunnels them to the tunnel exit-point using the specified tunnel parameters.

We provide an example of the tunnelling process here, and assume that a previous AR (PAR) is proxying for an MN's previous care-of address (PCoA), and the MN is away from the coverage area of the PAR.

When a CN sends packets to the MN's PCoA, the PAR intercepts such packets. The tunnelling protocol module (or TB) at the PAR then categorize the traffic into different traffic classes. Following by the classification, the TB at the PAR selects a corresponding tunnel with pre-configured parameters from its Forwarding Tunnel Table. And then the PAR encapsulates those original packets, and forwards the

packets to the new AR (NAR) through the specific tunnel. Upon receipt of the tunnelled packets, the TB at the NAR decapsulates the packets and verifies their traffic class, According to which, the NAR caches the packets using different buffer. At the same time, the NAR waits for the imminent MN for the delivery of the buffered packets.

2.5 Dynamic tunnelling key exchanging process

Each access router has some pairs of keys stored locally in its *Tunnelling Key Table.* An example of such table is shown in Table 2.4.

Table 2.4 an Example of Tunnelling Key Table

No.	Key Index	Public Key	Private Key	Partners	Valid	Traffic Class
1	K-P1	Pub-P1	Pri-P1	NAR	Yes	voice
2	K-P2	Pub-P2	Pri-P2	NAR	Yes	VoIP
3	K-P3	Pub-P3	Pri-P3	AR1	Yes	-
4	K-P4	Pub-P4	Pri-P4	AR2	No	VoIP

The key exchanging procedure happens when a tunnel end-point selects an entry from its Tunnelling Key Table, and generates a triplet of (key index, public key, and private key). Acting as the tunnel entry-point, the node (or AR) then sends the corresponding pair of (key index, public key) to the tunnel exit-point, which then performs the same operation.

When the tunnel entry-point wants to tunnel packets to the tunnel exit-point, it first selects a public key from a set of public keys. Such selection can be made upon different traffic class or node's preference. Then, the selected public key of the tunnel exit-point is utilized to encrypt the packets. And then, the tunnel entry-point sends the tunnelled packets and the key index of the corresponding public key to the tunnel exit-point. Upon receipt of the tunnelled packets, the tunnel exit-point first checks its Tunnelling Key table. According to the received key index, the tunnel exit-point finds the corresponding private key, using which the encrypted data are then decrypted. Figure 2.5 illustrates the tunnelling key exchanging and data encrypting procedure.

The advantages of the dynamic key exchanging approach are listed as follows:

- Each tunnel end-point has several pairs of (public key, private key);
- The concept of key index is introduced to distinguish each pairs of keys;

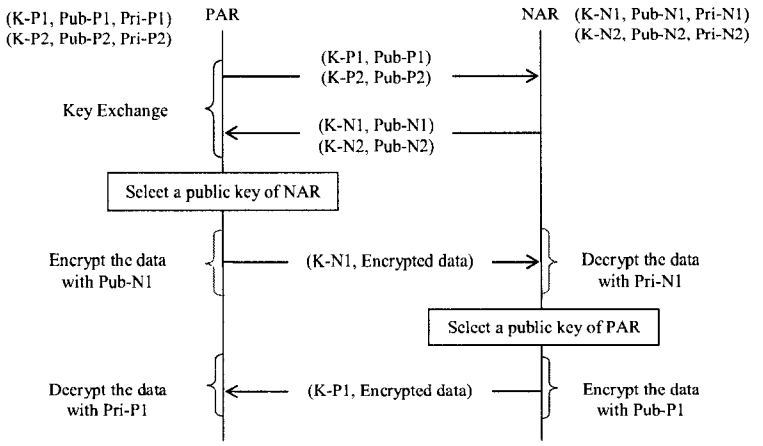

Figure 2.5 Tunnelling Key Exchanging Process

- Interception of the exchanged key index does not imply the knowledge of the corresponding public key if the key exchanging process is guaranteed to be secure enough.
- Selection of a public key from a set of keys can be based on different traffic class. For example, for delay-sensitive applications, we can choose a public key with fewer bits. This results in less processing overhead at tunnel end-point and lower processing time as well.

2.6 Key Generation and Utilization Process

We suppose that two tunnel end-points: Alice and Bob have obtained their own pair of (public key, private key) from a Public-Key Infrastructure (PKI), in which one or more third parties, known as Certificate Authorities (CAs), certify ownership of key pairs. The pair of keys are denoted as (K^+_{AP}, K^-_{AP}) for Alice, (K^+_{BP}, K^-_{BP}) for Bob.

Before beginning the communication with Bob, Alice asks for two session keys from the PKI. In response to this request, the PKI sends two session keys to Alice in the format: $K^+_{AP}(K_{AB}, K^+_{BP}(K_{AB}))$ and $K^+_{AP}(K_{ABI}, K^+_{BP}(K_{ABI}))$. K_{AB} will be utilized to secure the session between Alice and Bob while K_{ABI} for corresponding key index security.

Alice then decrypts the session keys with its private key with the PKI: K^-_{AP}. For the first message, Alice acquires K_{AB} and $K^+_{BP}(K_{AB})$. Then it goes to Bob's web site

to find Bob's public key with the PKI: K^+_{BP}. Alice then sends $K^+_{BP}(K^+_{BP}(K_{AB}))$ to Bob indicating that the session key K_{AB} will be utilized for a specific session, and will be updated or refreshed upon the termination of the session. For the second message from PKI, Alice obtains K_{ABI} and $K^+_{BP}(K_{ABI})$, it then sends $K^+_{BP}(K^+_{BP}(K_{ABI}))$ to Bob indicating that this key will be used for key index security of the specific session, and will be updated or refreshed upon the termination of the session.

Upon receiving these two messages, Bob decrypts twice with its private key with the PKI: K^-_{BP}, eventually gets the session keys: K_{AB} and K_{ABI}.

When Alice has a message M for Bob, it then first goes to Bob's web site, in which Bob published its own public key table as follows:

Table 2.5 Bob's Public Key Table

No.	Key Index	Public Key
0	BP	K^+_{BP}
1	B1	K^+_{B1}
2	B2	K^+_{B2}
3	B3	K^+_{B3}
4	B4	K^+_{B4}

Note that as Alice can obtain Bob's public keys and corresponding key indexes from Bob's web site, and we assume that Bob's web site is secure enough. As a result, the first tunneling key exchange procedure described in section 2.5 can be completely eliminated. In Table 2.5, the first record shows Bob's public key from the PKI, and the following records can be generated by Bob himself. By this means, the cost used to get those keys is minimized.

Alice then selects one public key from Bob's public key table. For example, Alice selects using Bob's public key K^+_{B1} to encrypt the message M, it then sends Bob an encrypted message in the format of $K_{AB}(K^+_{B1}(M), K_{ABI}(B1))$. Note that the corresponding key index $B1$ is encrypted using the session key K_{ABI}. Another session key K_{AB} is used to secure the whole session, which includes the encrypted message M and key index $B1$.

Bob then first decrypts this message with the session key K_{AB}, then it gets $K^+_{B1}(M)$ and $K_{ABI}(B1)$. Afterwards, Bob decrypts the second part with the session key K_{ABI}, then it acquires the key index $B1$. With the assistance of this key index, Bob finds the corresponding private key K^-_{B1} from its own key table. An

example of Bob's key table is shown in Table 2.6. Using the corresponding private key K_{B1}^{-}, Bob decrypts $K_{B1}^{+}(M)$. And finally, Bob obtains the message M. The key generation and utilization procedure is shown in Figure 2.6.

Table 2.6 Bob's Key Table

Key Index	Public Key	Private Key
BP	K_{BP}^{+}	K_{BP}^{-}
B1	K_{B1}^{+}	K_{B1}^{-}
B2	K_{B2}^{+}	K_{B2}^{-}
B3	K_{B3}^{+}	K_{B3}^{-}
B4	K_{B4}^{+}	K_{B4}^{-}

To be robust at the security, after selection of a public key from Bob's public key table, Alice can send Bob an encrypted message in the format as follows: $K_{AB}(K_{B1}^{+}(H(M)),\ M,\ K_{ABI}(B1))$. $H(M)$ denotes the hashed message. Note that as the Message-Digest algorithm 5 (MD5) is proven to be insecure, here we suggest using Secure Hash Algorithm (SHA) hash functions.

After receiving the encrypted message, Bob first decrypts using the session key K_{AB}, then it obtains $K_{B1}^{+}(H(M))$, M, $K_{ABI}(B1)$. Afterwards, Bob decrypts $K_{ABI}(B1)$ with the session key K_{ABI}, and gets $B1$. Bob then finds its corresponding private key: K_{B1}^{-} from its own key table with the help of $B1$. Using this key, Bob decrypts $K_{B1}^{+}(H(M))$, then acquires the value of $H(M)$.

Bob hashes the message M that obtained from the first decryption, and get $H(M)\_0$. Bob then compares $H(M)\_0$ with $H(M)$ which obtained from the second decryption. In case of difference, Bob then discards the message since it knows that someone knows the session key K_{AB}. Otherwise, the verification of data integrity is done with success. This robust key generation and utilization process is illustrated by Figure 2.7.

To be more robust for the security, after selection of a public key from Bob's public key table, Alice can send Bob an encrypted message in the format as follows: $K_{AB}(K_{B1}^{+}(H(M),\ Nonce),\ M,\ Nonce,\ K_{ABI}(B1, Nonce))$. Note that $H(M)$ denotes the hashed message while the $Nonce$ is used to avoid replay attack and the man-in-the-middle (MITM) attack.

After receiving the encrypted message, Bob first decrypts using the session key K_{AB}, then it obtains $K_{B1}^{+}(H(M),\ Nonce)$, M, $Nonce$, $K_{ABI}(B1, Nonce)$. Afterwards,

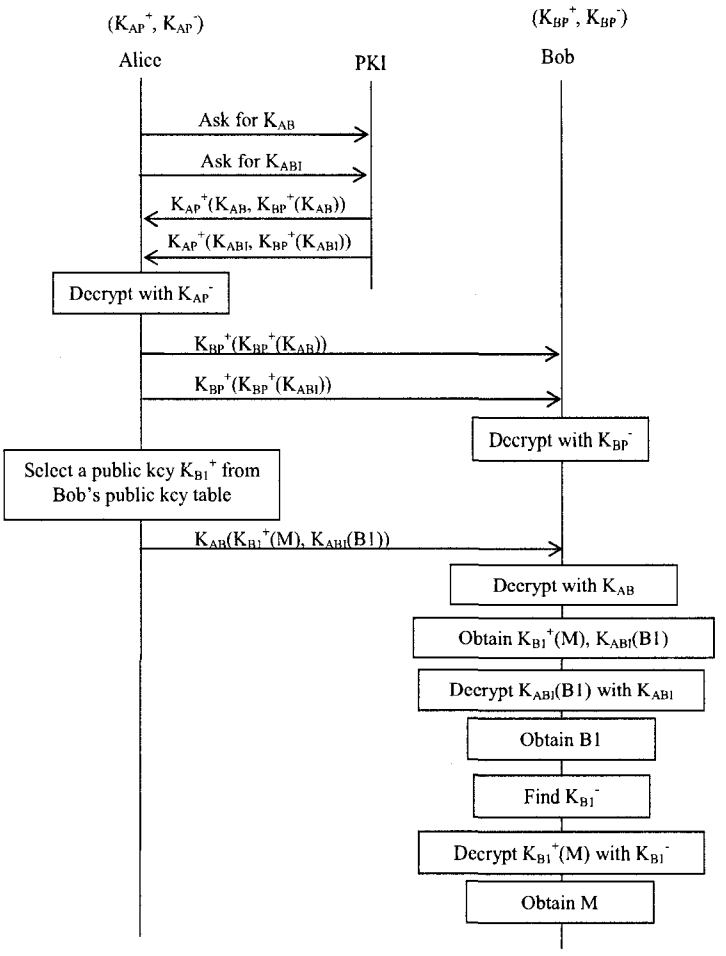

Figure 2.6 Key Generation and Utilization Procedure

Bob decrypts $K_{ABI}(B1, Nonce)$ with the session key K_{ABI}, it gets $B1$ and $Nonce$. Bob then compares the value of the nonce (number used once) obtained from the first decryption with that obtained from the second decryption. In case of difference, Bob then discards the message since it knows that someone knows the session key K_{AB}. Otherwise, Bob finds the corresponding private key K_{B1}^{-} from its key table with the help of the key index $B1$. It then uses this private key to decrypt $K_{B1}^{+}(H(M), Nonce)$. Bob gets the values of $H(M)$ and $Nonce$. Bob compares the value of this nonce with that obtained from the first decryption. In case of identical, Bob continues the verification for message integrity. To do so, Bob then hashes the message M obtained from the first decryption, and compares this value with $H(M)$ that obtained from the third decryption. In case of identical, Bob gets the original message M. Otherwise,

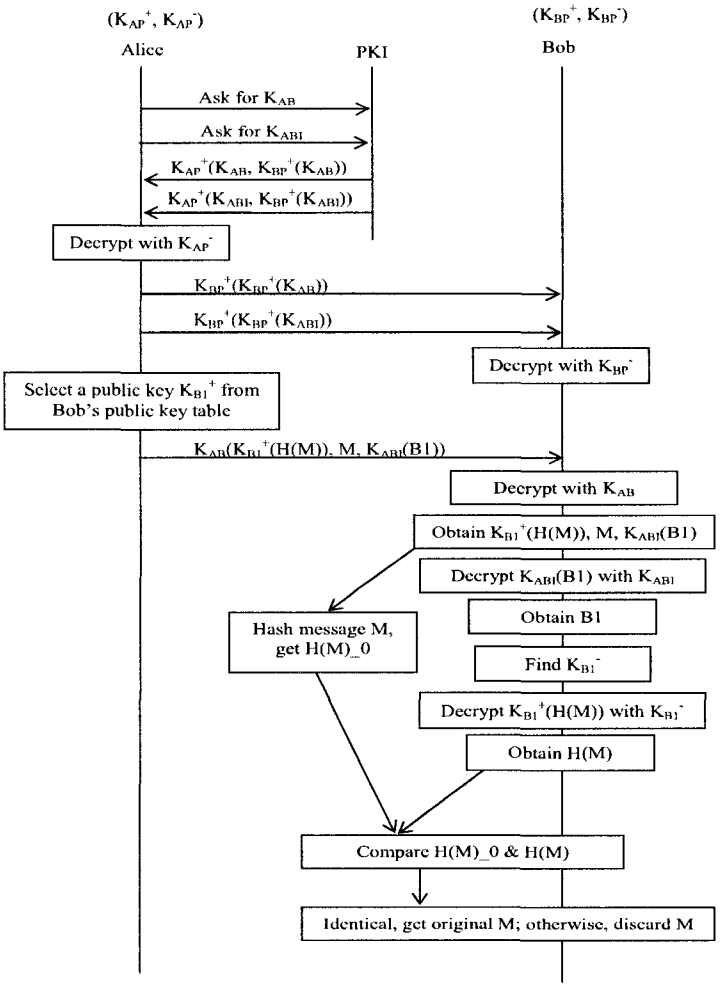

Figure 2.7 A Robust Key Generation and Utilization Procedure

Bob discards the message.

2.7 Conclusion

This chapter introduces a patent-pending solution, called access router tunnelling protocol. The proposed protocol is used to establish bidirectional secure tunnels between nodes in a network. These tunnels are provided with a set of minimal characteristics. Utilization of such tunnels can guarantee certain quality of service. In addition, new key establishment approach is designed, which allows access routers to possess different pairs of keys with respect to different traffic class or type. Moreover,

this chapter also proposes new key generation and utilization procedure, which allows either the sender or the receiver to have its own public key table and key table, and dynamically select public key from public key table. This procedure can also reduce the cost of obtaining keys from a third-party since it allows users to generate the keys by itself.

CHAPTER 3

PROPOSED INTEGRATED ARCHITECTURE FOR NEXT GENERATION WIRELESS NETWORKS

Advancement in wireless technologies enable mobile nodes (MNs) to benefit from different wireless networks such as wireless local area networks (WLANs), wireless metropolitan area networks (WMANs), third generation (3G) cellular networks, etc. These networks are expected to integrate with each other to provide ubiquitous and high data-rate services to roaming users [3]. However, integrating of such networks/systems brings about new challenges due to their heterogeneity. Under the circumstances, this chapter proposes a novel Architecture to support Fast authentication and Seamless roaming (AFS) in next generation heterogeneous wireless networks. AFS extends the existing infrastructure to integrate disparate wireless systems that include WLANs, WMANs and 3G systems using the technologies of universal mobile telecommunication system (UMTS) and cdma2000. Additionally, the designed architecture exploits IPv6 as interconnection protocol. In addition, new network elements such as eMAP and eHAAA are introduced into the novel architecture. As a result, the proposed architecture presents a unified infrastructure; this facilitates the deployment in the near future.

3.1 Introduction

So far, with the rapid progress in wireless networking and communications technology, mobile users have been capable to profit from various wireless systems, such as wireless personal area networks (WPANs) (e.g. Bluetooth), WLANs, WMANs, 3G wireless systems, etc. However, these systems present distinct characteristics in terms of radio access technology, bandwidth, delay, radio coverage area, installation cost, maintenance cost and quality of service (QoS) provisioning, etc. this makes intelligent and seamless integration of such systems a challenging issue in the design of next generation wireless networks (NGWNs).

The new integrated architecture for NGWNs must preserve the advantages of

individual system and eliminate the weaknesses of each network [3] while allowing mobile users to be always connected to the best available network [51]. Therefore, it should have the following features [3] [89]:

- **Economical**: Using as much existing infrastructure as possible, rather than putting more efforts into developing new radio interfaces and access technologies [2]. In the meantime, minimizing the use of new infrastructure is also important to ensure rapid deployment.

- **Scalable**: The new architecture should be able to integrate any number of wireless systems of same or different service providers.

- **Transparency to heterogeneous access technologies**: Underlying radio access technology should be transparent and imperceptible to mobile users.

- **Seamless mobility**: Network connectivity and session continuity should be guaranteed when mobile nodes (MNs) perform any kind of roaming, such as intra- and inter-technology handoffs.

- **Security**: Providing a level of security and privacy which is equivalent to or better than the existing wireless and wired networks is essential in the design of new architecture for NGWNs.

This chapter introduces a new Architecture to support Fast authentication and Seamless roaming (AFS) in NGWNs. The proposed AFS integrates 3G cellular data networks with WLANs, WMANs. Two major standards for 3G cellular networks: UMTS and cdma2000 are taken into account while designing the integrated architecture, since these standards are well-specified by the 3G partnership projects initiatives, i.e. 3GPP [61] and 3GPP2 [62].

Noreover, AFS integrates heterogeneous wireless systems using a new network entity called enhanced mobility anchor point (eMAP). Such node aggregates key functionalities of mobility anchor point (MAP) and domain gateway (wireless interworking gateway (WIG) for WLAN, gateway GPRS support node (GGSN) for UMTS, packet data serving node (PDSN) for cdma2000, and domain interworking gateway (DIG) for WMAN). It may also combine the functionality of local authentication, authorization and accounting (LAAA) server [90], and visitor location register (VLR) of the GSM

system. In addition, to realize seamless roaming, mobile service area is partitioned into regions, which are further divided into domains controlled by eMAPs.

Furthermore, AFS achieves transparency to heterogeneous access technologies by exploiting IPv6 [54] as the inter-connection protocol, making AFS presents an all-IP-based hierarchical structure augmented with new network entities: eMAPs and enhanced home AAA server (eHAAA). The eHAAA aggregates key functionalities of home AAA server (HAAA) [90] and a central database (CD) that contains detailed information about each region, e.g. region identifier, each eMAP's IP address and its neighborhood, regional edge node (REN) identifier, etc. The eHAAA may also contain the functionality of home location register (HLR).

3.2 Background and related work

Recently, low cost, high speed WLANs are expected to complement the 3G cellular networks for wireless Internet access in hotspot areas such as airports, train stations, hotels, restaurants and cafés. On the other hand, the complementary nature of 3G systems and WLAN has attracted industry, academia and standardization organizations for their integration [69]. However, the heterogeneities between 3G cellular networks and WLANs in terms of radio access technology, mobility management, security and privacy aspect, and QoS provisioning [70] bring many challenges to their interworking and integration.

In a meanwhile, the emergence of the worldwide interoperability for microwave access (WiMax) technology enables WMANs to provide high bandwidth, low-cost, scalable, real-time multimedia services to wireless users over long distances. However, 3G systems and WMANs present different characteristics such as data rate, coverage area, protocol and QoS aspect [71], integrating these two networks also becomes a challenging issue in NGWNs. A successful interworking 3G/WMAN architecture can provide wide-area coverage, strong mobility support and high-speed broadband wireless access for nomadic users.

In addition, integration of WLANs with WMANs presents a difficult issue as well [72]. The integrated WLAN/WMAN architecture may extend the coverage area of a WLAN and augment the service availability for mobile Internet applications in case where a wired infrastructure is unavailable, e.g. in remote rural or suburban areas.

So far, a number of solutions have been proposed to integrate heterogeneous wire-

less and mobile communication systems. Yet, most of research activities focus on design architecture for integrating 3G systems with WLANs [3], [69], [70], [73]-[85]. Amongst all, 3GPP has introduced six scenarios for interworking between UMTS and WLAN and discussions about the functionality, architecture and feasibility of such interworking are still working in progress [73], [74]. In addition, interworking between WLAN and cdma2000 are well addressed by 3GPP2 in [75], [76]. However, only a few solutions are designed for integration of WLANs with WMANs [72], and for interworking 3G systems with WMANs [71], [86].

Effectively combining 3G systems and WLANs into an integrated wireless data access environment empowers mobile users with ubiquitous connectivity and high-speed services, especially in hotspot areas where bandwidth-sensitive applications are most demanded. And interworking between WLANs and 3G systems can be implemented in different ways. The WLAN could be an integral part of the 3GPP system or the two systems could be separating [73], [74]. The way of integration can be classified into *tight coupling* [74], [77], [84], [85], [87], *loose coupling* [78], [87], *no coupling* [79], [80], [88] and *hybrid coupling* [83].

Tight coupling (or *emulator approach*) allows a WLAN to appear to the 3G core network as either a radio access network (RAN) in case of general packet radio service (GPRS) [74], [77] or as a packet control function (PCF) in case of cdma2000 [84]. Using tight coupling, the cellular radio is simply replaced by the WLAN radio providing equivalent functions in a RAN. Hence, it enables the reuse of 3G system protocols and existing network infrastructures. Moreover, tight-coupled approach enables WLAN user to access 3G system services with guaranteed QoS support and seamless mobility [83]. However, a new interface is needed between the 3G core network and a WLAN, which implies that WLAN and 3G system must belong to the same operator. Consequently, independently operated WLANs cannot be integrated with 3G networks [79], [81], [87], thus results in less flexibility of integration. Furthermore, MNs must implement the corresponding 3G protocol stack on top of the standard 802.11 [87], this adds design complexity on the user side. Moreover, as all the packets from incorporated WLANs go through the 3G core network, integrating points such as GGSNs become traffic bottlenecks [79], hard to accommodate bulky traffic. To fix this, 3G network elements such as PDSNs and GGSNs must be modified or upgraded to sustain the increased load from each integrated WLAN.

Loose coupling (or *mobile IP approach*) allows a WLAN to connect to the 3G core

network indirectly via an external IP network such as the Internet [70], [87]. With loose coupling, the data paths of WLANs are completely separated from those of the 3G systems [87]. And different protocols are adopted to handle authentication, mobility management and billing. However, to achieve seamless integration, these protocols must inter-operate with each other. For example, in case of cdma2000/WLAN, WLAN must support mobile IP and proxy AAA (P-AAA) functionalities for mobility management and billing issues. Moreover, loose coupling allows independent deployment and traffic engineering of WLAN and 3G systems. And with roaming agreements with many providers, MNs can obtain all network access from a sole service provider. Moreover, loose-coupled architecture also provides opportunities for wireless Internet service providers (WISPs) to extend their service scopes/areas via roaming agreements with other WLANs and 3G operators. The major weakness of this approach is excessive handoff latency, because mobility signaling traverses a relatively long path due to the separation between 3G RAN and WLAN domains [70]. And high handoff delay leads to unacceptable packet losses, traffic congestion and handoff failure, this makes loose coupling unable to support service continuity during inter-system handoff [83]. Nonetheless, loose-coupled architecture is the most preferred solution for seamless integration of WLAN and 3G networks [70], [87].

No coupling (or *gateway approach*) treats WLAN and 3G system as peer-to-peer networks [69]. The principal idea of this approach is exploiting legacy mobility management protocols to handle intra-system (or horizontal) handoff and using roaming agreements to handle inter-system (or vertical) handoff via a logical node, called interworking gateway (IG) [3] or interworking decision engine (IDE) [89]. The gateway exchanges necessary information between WLAN and 3G system, converts signals and routes packets for roaming users [79], [80], [88]. The advantage of this approach is that the integrating networks can operate independently. And unlike loose coupling approach, mobile IP functionalities are unnecessary, thus results in less handoff delays and packet losses [79].

Hybrid coupling architecture integrates WLAN with UMTS and allows the UMTS core network to efficiently accommodate traffic from the integrated WLAN. Data paths are differentiated according to traffic type. For real-time traffic, tight coupling approach is exploited, so WLAN data traffic traverses the UMTS core network. For non-real-time traffic, loose coupling is necessary to force WLAN data packets go through an external IP network. Furthermore, an access point gateway (APGW) is

introduced to connect an access point (AP) in the WLAN with the serving GPRS support node (SGSN) in a UMTS terrestrial radio access network (UTRAN), and new functionalities such as traffic differentiation and setup data path are added into the APGW [83].

3.3 Proposed integrated architecture

There are two possible solutions in design integrated and interworking architecture for NGWNs: either developing novel wireless systems with radio interfaces and technologies to satisfy the requirements of the services demanded by future mobile users or integrating the existing wireless systems [3]. Since the former solution is expensive and impractical, we advocate using as much existing architectures as possible, same idea presented in [3], [89].

On the other hand, unlike the architecture for ubiquitous mobile communications (AMC) [3] and the integrated intersystem architecture (IISA) [89], our proposed architecture for NGWNs does not imply pre-existence of roaming agreements (RAs) or service level agreements (SLAs) between each pair of participants (or service providers). Nevertheless, mobile users can obtain services in a new visiting domain either via direct RA or through an indirect but trusted third-party. The new architecture to support fast authentication and seamless roaming (AFS) integrates disparate wireless networks such as WLAN, WMAN and 3G systems that include UMTS and cdma2000 networks, shown in Figure 3.1.

Mobile service areas are partitioned into regions, which further divided into domains using heterogeneous or homogeneous radio access technologies. These domains overlap each other and formulate overlaid multitier networks. Each domain is managed by a new network element, eMAP. Such node combines key functionalities of mobility anchor point (MAP), which is defined in hierarchical mobile IPv6 (HMIPv6) for mobility management [8] [9], and domain gateway (WIG for WLAN, GGSN for UMTS, PDSN for cdma2000, DIG for WMAN). It is also possible for eMAP to aggregate the functionality of local authentication, authorization and accounting (LAAA) server [90] and visitor location register (VLR). Note that the generic functionality of domain gateway is to perform IP network address and port translation (NAPT) to enable access network operator to use private-space IP addresses inside its domain as well as simultaneously provide external IP network access. In addition, using the

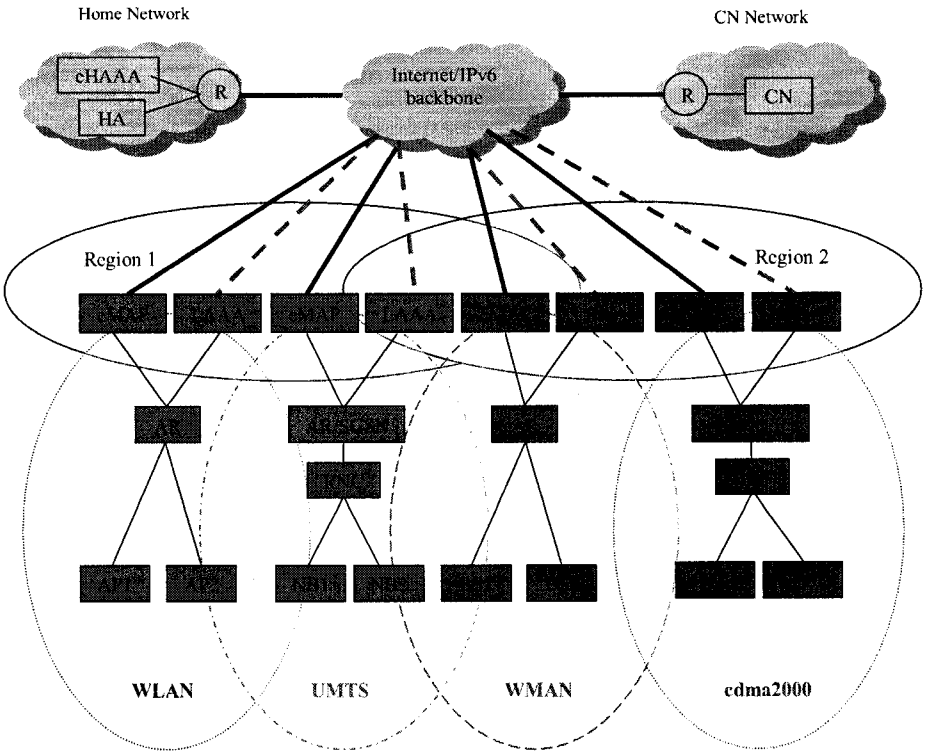

Figure 3.1 Proposed Integrated Architecture for NGWNs

concept of eMAP, AFS presents a unified infrastructure, this facilitates real deployment.

Another new network entity, called enhanced home AAA (eHAAA) server, is also introduced in AFS. It aggregates key functionalities of home AAA server (HAAA) [90] and a central database (CD) that contains detailed information about each region, e.g. region identifier, each eMAP's IP address and its neighborhood, regional edge node (REN) identifier, etc. The eHAAA may also contain the functionality of home location register (HLR).

Solutions currently found in the literature attempt to reduce authentication delays in a single administrative domain, where centralized control is preferred [169]. However, as the demands for seamless roaming across different wireless systems (or domains) increase, new authentication solution is required for inter-domain (or inter-system) handoff. Since each domain is under its own authority and administration [169], and each authority exploits its own authentication mechanism that is unavailable to others. This situation makes third-party-based authentication solution turn

out to be infeasible. In this context, we propose trust-based fast authentication scheme to replace third-party-based solution. Before elaborating the proposed fast authentication mechanism, we develop a new network selection mechanism that allows MNs to intelligently and efficiently select the most appropriate network (or the best connected network) during handoff.

3.4 New Network Selection Method for AFS

Entering into a new administrative domain, an MN needs to select the best connected network before breaking its current connection. Therefore, it performs the following network selection procedures:

1. Analyze wireless environment within range by listening to broadcast channels and examining the received broadcasting messages.

2. Select a network with the same radio access technology (RAT) and having direct roaming agreement (RA).

3. Otherwise, select a network with the same RAT and indirect RA. The latter means the concerned network operators can establish an indirect temporary virtual RA via a trusted third-party.

4. Select a network with integration capability and direct RA with the MN's home network provider.

5. Otherwise, select a network with integration capability and indirect RA via a trusted third-party.

The main idea is to avoid vertical handoff as much as possible, if applicable. Otherwise, select a network with direct RA; If the former two requirements cannot be met, then select a network with which the MN's home service provider could establish a temporary virtual indirect RA via a trusted third-party.

To facilitate network selection on the user's side, AP in WLANs inserts some necessary information into their broadcast *Beacons*. Besides the Service Set Identifier (SSID), the AP may broadcast the information such as WLAN_NAME, integration capability, list of 3G partner_NAME (or WMAN partner_NAME) and Mobile Country

Code (e.g. butterfly: integration capability/yes: 3G_Telus: 001). This enables MNs to make intelligent decision about network selection.

The same principle is applicable for WMANs. In this case, Base Station (BS) of WMANs broadcasts information about its network_NAME, integration capability, list of 3G partner_NAME (or WLAN partner_NAME) and Mobile Country Code. Such information assists MNs to select a trusted domain and handoff to that domain with which its own home network can establish a temporary RA via a trusted third-party. By this means, after successful authentication, all outgoing data traffic from the MN will be tunneled to a neighboring trusted domain, then forwarded to the destination node. At the same time, all incoming traffic will be intercepted by a previous trusted domain and tunneled to the new domain, then decapsulated and forwarded to the MN.

The same principle is also applicable for 3G systems. In this case, Node B (NB) of UTRAN or Base Transceiver Station (BTS) of cdma2000 access network broadcasts information about its network_NAME, integration capability, list of WLAN partner_NAME (or WMAN partner_NAME) and Mobile Country Code. Such information assists MNs to select a trusted network and handoff to that network with which its own eHAAA can establish a temporary RA via a trusted third-party: eMAP.

To facilitate network selection, the MN should have preknowledge about the partners of its home service provider. For example, assuming that the home service provider of the MN is A, and A's trusted partners include B and D. When this MN moves into a new visiting domain controlled by service provider C and receives beacons from APs with the following information:

- *butterfly*: integration capability/yes: $3G\_D$: 001
- *supercool*, integration capability/yes: $WMAN\_F$: 001

Then the MN may select the WLAN whose name is *butterfly* because it is sure that an indirect temporary RA will be established between A and the operator of the *butterfly*: C, via D.

3.5 Proposed Fast Authentication and Seamless Roaming Schemes for AFS

To achieve fast authentication, mobile service area is partitioned into regions, which are further divided into domains controlled by an eMAP. And the eHAAA manages a Central Database (CD) about details of each region, e.g. Region Identifier (Reg_ID), each trusted eMAP's IP address and its trusted neighborhood, Regional Edge Node (REN) identifier as well. Such a database allows eHAAA to quickly multicast an MN's AAA information to corresponding trusted eMAPs upon receiving a Binding Update (BU) message from the MN. Hence, authentication related materials are always one-hop ahead of MN's current associated eMAP. As a result, once the MN moves into a new network, end-to-end authentication is replaced by local authentication with the new selected trusted eMAP, this results in reduced authentication latency during handoff.

Furthermore, in case where the MN moves into a region without trusted eMAP, a randomly selected eMAP may request its trusted neighboring eMAPs for the MN's authentication materials, thus a temporary virtual trust relationship can be established via a third-party.

Generally, The main idea of trust-based fast authentication scheme is that neighboring eMAPs that trust each other share authentication material (e.g. security key) for a visiting MN to avoid lengthy authentication routines each time the MN switches mobility domains. To compensate for trust relationship hole, i.e. the new selected eMAP has no trust relationship with the MN's home service provider, the new eMAP may request for the MN's authentication material from its trusted neighbors.

When entering into a new eMAP domain, an MN receives *Router Advertisements* from Access Routers within range. Then, it selects a trusted eMAP domain using the above-mentioned network selection technique, and configures two Care-of Addresses (CoAs): a Regional CoA (RCoA) on the selected eMAP's link and an on-link Local CoA (LCoA). Both CoAs can be formulated in a stateless manner [43]. After generating the CoAs, the MN sends a *Local Binding Update (LBU)* to the eMAP. Note that here eMAP works as Mobility Anchor Point (MAP) of Hierarchical Mobile IPv6 (HMIPv6) [8].

Upon receiving the LBU, the eMAP searches its address pool to verify the uniqueness of the MN's RCoA. Generally, HMIPv6 requires the MAP to perform Duplicate

Address Detection (DAD) process for the RCoA. However, the DAD takes about one second to be completed [170]. In this context, we propose utilizing an address pool at the eMAP to reduce the latency caused by DAD process. After confirming that the RCoA is unique, the eMAP binds the MN's RCoA to its LCoA and returns a *Binding Acknowledgement (BA)* to the MN. Following a successful registration with the eMAP, a bidirectional tunnel is established between the MN and the eMAP.

After local registration with the new eMAP (neMAP), the MN sends a BU to its home agent, which then forward this BU to the eHAAA server. This BU will assist the HA to bind the MN's Home Address (HoA) with its RCoA. Note that to reduce the latency between the HA and eHAAA server, their functionalities can be combined together in a manner that formulates a new network element, called home intelligent node (HIN).

The eHAAA then extracts the network prefix from the received RCoA and searches the neMAP's IP address from its central database. Subsequently, the eHAAA verifies whether the neMAP is a Regional Edge Node (REN) or not. If so, the MN's authentication materials will be forwarded to all trusted eMAPs located in the same region as the neMAP as well as those trusted eMAP in the neMAP's neighboring regions. Otherwise, the eHAAA only needs to send the MN's authentication materials to all trusted eMAPs located in the same region as the neMAP. Therefore, when the MN performs inter-domain movements, it only needs to send an *Authentication Request* to the local eMAP, which in turn verifies its Authentication Database (AuD) to authenticate the MN. After successful confirmation, the eMAP replies the MN with an *Authentication Response.* By this means, authentication for successive handoff is carried out locally instead of in an end-to-end way.

In case where the associated neMAP has no preknowledge about the requested MN's authentication materials, the eMAP may multicast a *Further Authentication Request* to its trusted neighborhood. Then those eMAPs who store the corresponding authentication materials will forward them to the neMAP, which then replies the MN with an *Authentication Response.*

In addition, trusted eMAP can also distribute the authentication materials of an MN to all the access routers (ARs) in its mobility domain. In this case, authentication can be carried out between the MN and a new AR instead of between the MN and eMAP. This allows further reduction of authentication delays during handoff.

After entering into a new AR's coverage area, the MN may send an *Authentica-*

tion Request to the AR. If the AR has the authentication materials of the MN, it simply replies with an *Authentication Response.* Otherwise, the AR may request the associated eMAP or its neighboring ARs for the authentication materials by sending *Further Authentication Request* message. Those adjacent ARs or the associated eMAP can reply with a *Further Authentication Response.* As a result, the new AR forwards an *Authentication Response* to the MN indicating a successful completion of authentication.

Figure 3.2 illustrates an example of fast authentication scheme. When an MN first logs into the system, its related authentication materials are distributed by the eHAAA to corresponding eMAPs after successful home registration. For example, an MN enters into Region 1 and associates with eMAP_1, the eHAAA multicasts the MN related authentication materials to eMAP_3 and eMAP_5. In case where the MN first associates with eMAP_5, its corresponding authentication materials are distributed by the eHAAA to eMAP_1, eMAP_3, eMAP_5, eMAP_6 and eMAP_9. Note that each eMAP represents an administration domain, either with homogeneous access technology or with heterogeneous radio technology. In case where the MN needs to connect to eMAP_4, it sends an *Authentication Request* to eMAP_4, which then multicasts a *Further Authentication Request* to its neighborhood: eMAP_3 and eMAP_5. Either eMAP_3 or eMAP_5 may sends an *Further Authentication Response* to eMAP_4. Such *Further Authentication Response* should include the MN's authentication materials. After successful authentication, the eMAP_4 sends an *Authentication Response* to the MN.

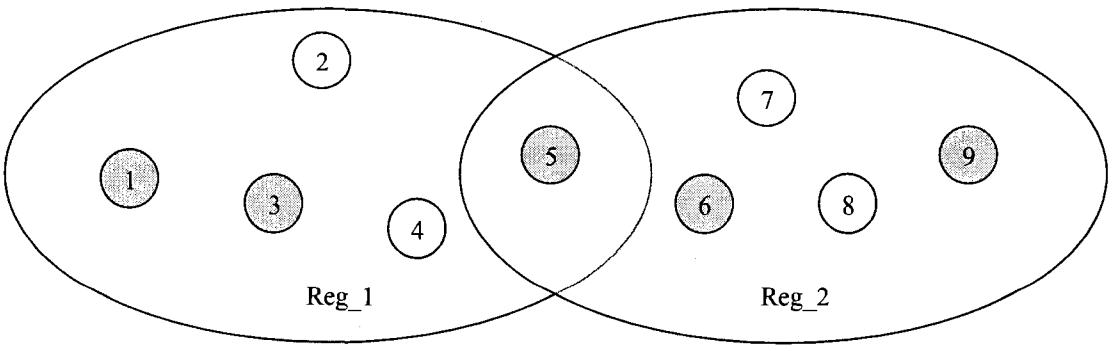

Figure 3.2 An Example of Fast Authentication Scheme

To facilitate billing management, trusted eMAPs may send billing reports directly

to the eHAAA. However, for those eMAPs with temporary connection, the outgoing traffic from the MN will always be tunneled to a trusted eMAP, which then generates the billing report. And the incoming traffic from the Correspondent Node (CN) will be intercepted by the previous trusted eMAP, tunneled to the neMAP and forwarded to the MN. In both case, billing reports are generated by trusted eMAP.

In contrast to the most commonly recommended third-party-based authentication [3], [168], [69], our proposed authentication exploits trust-based relationship and replaces end-to-end authentication with local authentication. In addition, to remedy the case where there is no trust relationship between the home service provider and the visiting domain, a temporary virtual trust relationship is established via a trusted third-party.

3.6 Characteristic Features of AFS

In this section we qualitatively evaluate our proposed architecture, AFS, in the context of the design objectives, its advantages and disadvantages.

- **Economical**: AFS uses the existing infrastructure of wireless systems. It does not imply any change to the infrastructure of integrated wireless networks. AFS achieves seamless integration of heterogeneous wireless networks by using the eMAP, which extends the functionalities of Mobility Anchor Point (MAP) by adding new functionalities, such as authentication unit, roaming agreement management unit, billing management unit, trust relationship database, etc.

- **Scalability**: AFS can integrate any number of exiting or future wireless systems of same or different operators. In addition, these operators may not have direct roaming agreements or service level agreements among them. In addition, AFS does not imply any pre-existing roaming agreements between each pair of participants, thus it is completely scalable.

- **Transparency to heterogeneous access technologies**: AFS utilizes IPv6 as the common interconnection protocol. Hence, this IP layer based solution allows mobile nodes to roam among multiple wireless systems, regardless to different radio access technologies.

- **Security**: AFS can adopt existing security mechanisms to provide security and privacy equivalent to existing wireless networks.
- **Seamless mobility**: AFS supports seamless intra- and inter-system mobility using hierarchical mobile IPv6 (HMIPv6) as the mobility management protocol. Furthermore, it allows mobile node to intelligently select new access network to avoid vertical handoff as much as possible.
- **Fast authentication**: AFS supports fast authentication by using eHAAA and eMAP. Instead of end-to-end authentication, eHAAA distributes authentication materials of an MN to trusted eMAPs before the node moves to the new eMAP domain. Hence, authentication is always done locally. As a result, the delay concerned to the authentication procedure can be reduced significantly. In addition, the functionality of the eHAAA can be integrated into the Home Agent, to formulate a home intelligent node (HIN).

The limitation of the proposed architecture is that mobile nodes have to be modified in order to be equipped with the intelligence of network selection. This can add design complexity on mobile device's side. Moreover, eHAAA must have a global vision of regions, domains. In other words, it should have pre-knowledge about trusted eMAP's IP address, their neighborhood. This adds new complexity at the traditional HAAA server. Furthermore, multicasting authentication materials of an MN to trusted eMAPs may lead to high signaling costs. However, as authentication is combined with mobility management, handoff delays can be significantly reduced. In addition, as authentication is carried out locally either between the MN and eMAP or between the MN and AR, authentication latencies are largely minimized. And with the proposed network selection method, MNs can intelligently select the best connected wireless network and avoid vertical handoff as much as possible.

3.7 Conclusion

This chapter proposes a new integrated architecture for next generation heterogeneous wireless networks. Such interworking architecture presents the features such as scalability, adequate security and privacy, using IPv6 as interconnection protocol to

hide the heterogeneity of different wireless systems as well as the radio access technology. Moreover, new network selection, fast authentication and seamless mobility mechanisms are designed for the AFS.

Since authentication and mobility management procedures are combined together, the delays concerned about these two procedures can be minimized during handoff. Additionally, new seamless mobility scheme, such as Seamless Mobile IPv6 (SMIPv6) which will be elaborated in the following section: Section 4.4, can be used for the AFS. In order to efficiently combine with fast authentication scheme, authentication materials of any mobile node can be further distributed to access router by an eMAP. By this means, authentication can be done more quickly than using HMIPv6 for mobility management. Additionally, there will no security hole in this case as eMAP manages all the access routers in its own domain. In the near future, we plan to evaluate the performance of the AFS by using analytical models and simulations.

CHAPTER 4

PROPOSED SEAMLESS HANDOFF SCHEMES IN NEXT GENERATION WIRELESS NETWORKS

The commercial proliferation of multi-mode wireless terminals (such as laptops, personal digital assistants (PDAs), smart-phones), the explosion of mobile data communications, the emergence of multi-technology environments with disparate capabilities, the integration of such diverse wireless environments and the wide variety of traditional and value added services (VAS) offered to end-users have placed extra requirements for mobile communications and computing.

On the other hand, traditional voice, fax, email and paging services are increasingly being substituted by real-time multimedia applications, such as video-conference, image transfer and audio-streaming. However, the provision of such multimedia services imposes strict quality of service (QoS) requirements on the networks. Furthermore, since next-generation wireless networks (NGWNs) must not only provide integrated services, but also capabilities for dynamically relocating mobile terminals [31], new challenges have sparked research laboratories to realize an advanced level of tetherless, seamless multimedia services for mobile users roaming between different domains or systems. Given such circumstances, mobility management becomes an important issue to enable telecommunication networks to locate roaming terminals and deliver calls or data to them while these terminals moving into a new service area.

Currently, assorted wireless technologies and networks exist to meet the various needs of mobile users. For example, wireless local area networks (WLANs) enable the delivery of high data-rate services with small radio coverage, cellular networks provide voice and data services with a relatively lower data-rate, yet large coverage, and satellite networks enable global roaming with worldwide coverage at drastic costs. Since these networks are designed for specific service requirements and vary significantly in terms of bandwidth, delay, coverage area, costs and QoS provisioning [3], they tend to complement one another, and their integration enables mobile users to be "always best connected" [51] to the most appropriate network.

On other hand, NGWNs introduce a variety of heterogeneities in terms of radio access technologies, network architectures, network protocols and service demands, and such inherent heterogeneities require a common infrastructure to interconnect multiple access systems [52]. Hence, using all-IP-based infrastructure to support ubiquitous communication appears to be very promising. First of all, IP-based wireless networks are better suited to support the rapidly growing mobile data and multimedia applications [53]. This is confirmed by the fact that IPv6 [54] is designated as the only IP version supported for IP multimedia subsystem (IMS) within the third generation partnership project (3GPP) [55]. Secondly, IP-based wireless networks have already brought global success to Internet services and they will also prove to be a successful platform to foster future mobile services [53]. Last but not least, IP-based wireless networks are independent of the underlying radio access technologies, making it possible and feasible to maintain seamless connectivity over different radio technologies, while offering global roaming capabilities [53]. Therefore, NGWNs are designed to take advantage of IP-based technologies to achieve global roaming amongst a variety of access technologies [52].

This chapter first provides a survey of IP-layer mobility management protocols designed within the working groups of the Internet Engineering Task Force (IETF) for IPv6-based NGWNs. The remainder of the chapter is organized as follows. The challenges for mobility management in NGWNs are outlined in order to provide a global view of the research background. Typical IP-layer mobility management protocols such as mobile IPv6 (MIPv6) [7], hierarchical mobile IPv6 (HMIPv6) [8] [9], and fast handovers for mobile IPv6 (FMIPv6) [10] [11], fast handover for hierarchical mobile IPv6 (F-HMIPv6) [16]-[19] and proxy mobile IPv6 (PMIPv6) [20] are described in detail. Future trends in the design of intelligent mobility management protocols are then presented. Then we proposed seamless handoff schemes (SMIPv6) for mobile IPv6 (MIPv6)-based next-generation wireless networks.

4.1 Basic concepts and definitions

Generally, mobility management comprises two components: location management and handoff management [32]. In NGWNs, mobile users perform two types of movements: intra- and inter-system roaming. Intra-system roaming refers to movements between the cells within a system, and its mobility management solutions are based

on similar network interfaces and protocols. On the other hand, inter-system roaming refers to movements between different technologies, protocols, backbones or service providers. Based on intra- or inter-system roaming, the corresponding location management and handoff management can be further classified into intra- and inter-system location management and handoff management [52].

Location management is the process that allows the network to locate the access point (AP) of the mobile node (MN) for calls or data delivery. This process can be further split into two steps: location update and call (or data) delivery. The former requires mobile users to provide the network with their location information, while the latter indicates that the network is queried for the location information of mobile users in order to deliver calls (or data) to them. The challenges in the design of inter-system location management protocols include [52]:

- How to reduce signalling loads and latencies that pertain to service delivery.
- How to guarantee on-demand QoS in different wireless networks (or systems).
- How to select the most suitable network for mobile users to perform location registrations (or updates) when the service areas of heterogeneous wireless networks fully overlap.
- Where and how to store the mobile user's updated location information when the service areas of heterogeneous wireless networks fully overlap.
- How to determine mobile users' exact location within a specific time constraint when the service areas of heterogeneous wireless networks fully overlap.

Typically, handoff (or handover) refers to the process of transferring signals from one AP to another. Handoff management aims to maintain network connectivity as mobile users change their points of attachment to the network. Obviously, handoff protocols need to preserve connectivity as mobile users move about, while simultaneously curtailing disturbance, or disruption from ongoing call (or data sessions) transfers. Therefore, minimal handoff disruption and session continuity are the primary goals of handoff management [56]. Generally handoffs require certain key features such as low latency, minimal packet loss, minimal jitter for multimedia streaming, high reliability, sustainable scalability to large networks, etc. Here are some of the challenges encountered when designing handoff management protocols [52]:

- Minimizing signalling overheads and power consumption related to handoff management;
- Making efficient use of network resources during handoff;
- Improving network scalability, reliability and robustness;
- Where and how to store the mobile user's updated location information when the service areas of heterogeneous wireless networks fully overlap.
- Guaranteeing on-demand QoS during handoff: (1) reducing intra- and inter-system handoff latency which is composed of signalling message processing time, resources allocation and route setup delay, format transformation time, etc. (2) alleviating user-perceptible service degradation (3) decreasing handoff failure to a near-zero level and (4) mitigating packet loss rate to a near-zero level to achieve seamless mobility

4.2 Overview of IP layer mobility management protocols

In all-IP-based NGWNs, mobile nodes (MNs) freely change their APs while communicating with correspondent nodes (CNs). Accordingly, mobility management consists of a critical issue, which is to track mobile users' current location and to efficiently route packets to them. Several typical mobility management protocols are designed within the IETF working groups, such as MIPv6, HMIPv6, FMIPv6, F-HMIPv6 and PMIPv6. Moreover, a number of working groups are striving to improve the performance of such specifications.

Generally, IP mobility includes macro- and micro-mobility. Macro-mobility designates mobility over a large area; this refers to situations where MNs move between IP domains [34]. Typically, mobile IPv4 (MIPv4) [5] [6] and MIPv6 [7] are best suited for macro-mobility management. Nevertheless, this chapter only addresses mobile IPv6. Micro-mobility refers to mobility over a small area, i.e. within an IP domain. Usually, micro-mobility protocols confine movements related signalling within a local domain without propagating to the mobile user' home network and those hosting their communicating peers. Location updates to a distant home agent (HA)

and CNs are eliminated as long as MNs remain inside their local domain. Hence, micro-mobility protocols yield better performance than macro-mobility solutions for intra-domain roaming. This section presents macro-mobility protocol, like MIPv6 and micro-mobility protocols, such as HMIPv6, FMIPv6, F-HMIPv6 and PMIPv6.

4.2.1 Mobile IPv6 (MIPv6)

Mobile IPv6 (MIPv6), IETF's mip6 work group's baseline, was designed for macro-mobility management protocols for future all-IP wireless networks [7]. It enables MNs to remain connected while moving around within the Internet topology. An MN is always identified by its home address, regardless of its current location on the Internet. While away from the home network, the MN is associated with a care-of address (CoA). To route packets to its current location in an efficient manner, binding between an MN's home address and the CoA is managed by a home agent (HA), which is actually a router on the MN's home link. Packets destined for the MN are intercepted by the HA and tunnelled to the MN's current CoA. Consequently, given this tunnel mode, triangular routing problems become unavoidable. Furthermore, HAs also create bottlenecks in the network. To avoid triangular routing, MIPv6 defines route optimization that enables an MN to register its current location with a CN. As a result, the CN can directly send packets to the MN without passing though the HA.

MIPv6 mobility management procedure consists of movement detection, new CoA configuration, duplicate address detection (DAD) and binding registration (or update) with the HA and all CN with route optimization mode. Movement detection aims to determine whether the MN has moved to a new network or not. Such detection is important since, in situations where the MN has moved, the addresses configured in a previous network become invalid, requiring additional configuration to establish or maintain upper layer connectivity. Generally, this is accomplished by assessing the reachability of the current (or default) router, the validity of the configured addresses and finding a new available router on the network. To accelerate movement detection, MNs can multicast router solicitation (RS) messages requesting immediate router advertisements (RAs) from access routers, rather than wait for the subsequent periodic advertisements to arrive [42].

Based on the RA it received, the MN then configures a CoA on the new link.

The address can be configured in two different manners. First, using a stateless way, by combining the MN's respective interface identifier with the network prefix found in the RA [43], or it can occur via a stateful way during which a server assigns an address using the dynamic host configuration protocol (DHCP) [46].

When the new address is added to the specific interface, detection on duplicate address is performed in order to preserve the uniqueness of the new address. Hence, the MN sends a multicasting neighbor solicitation (NS) to its neighbors and this procedure requires waiting at least 1 second to listen to responses from other nodes. If no responses are received before the time limit expires, the new IPv6 address is considered available [42]. To alleviate this lengthy delay, optimistic DAD procedure [57] is preferred for fast address configuration.

Home registration involves the exchange of signalling messages (binding update (BU) and binding acknowledgment (BAck)) between the MN and the HA to enable the creation and maintenance of a binding between the MN's home address and its CoA. Accordingly, packets destined for the MN will be intercepted by the HA and tunnelled to its CoA. Such registration is authorized by the use of IPsec [58].

In case of route optimization (RO) that follows a successful home registration, correspondent registration is executed via a BU message sent by the MN in order to create the same binding for the CN. However, as empowering nodes with the ability to redirect packets from one IP address to another raises security concerns, return routability (RR) is performed prior to the correspondent registration.

The return routability procedure consists of home address test and CoA-test. During the home address test, the MN sends a *home test init* (HoTI) message to the CN via the HA. The CN responds with a *home test* (HoT) message containing a secret home keygen token. Addressed to the MN's home address, the home test message is forwarded to the MN's current CoA by the HA. While CoA-test is in progress, the MN sends a *care-of test init* (CoTI) message to the CN, which returns a *care-of test* (CoT) message that contains a secret care-of keygen token. The care-of test init and care-of test messages are directly routed between the MN and the CN, bypassing the HA.

The MN needs both the home keygen token and the care-of keygen token to validate the ensuing correspondent registration. The former token proves the MN to be the legitimate owner of its home address, while the latter token confirms that the MN is currently present at the new CoA. The return routability procedure also

enables the CN to be reasonably confident that the MN can be reached at both its CoA and its home address. Afterwards, the MN sends a *BU* message to the CN through a direct path. Accordingly, the CN sends packets directly to the MN's CoA without involving the HA. Figure 4.1 illustrates the mobility management procedure in MIPv6.

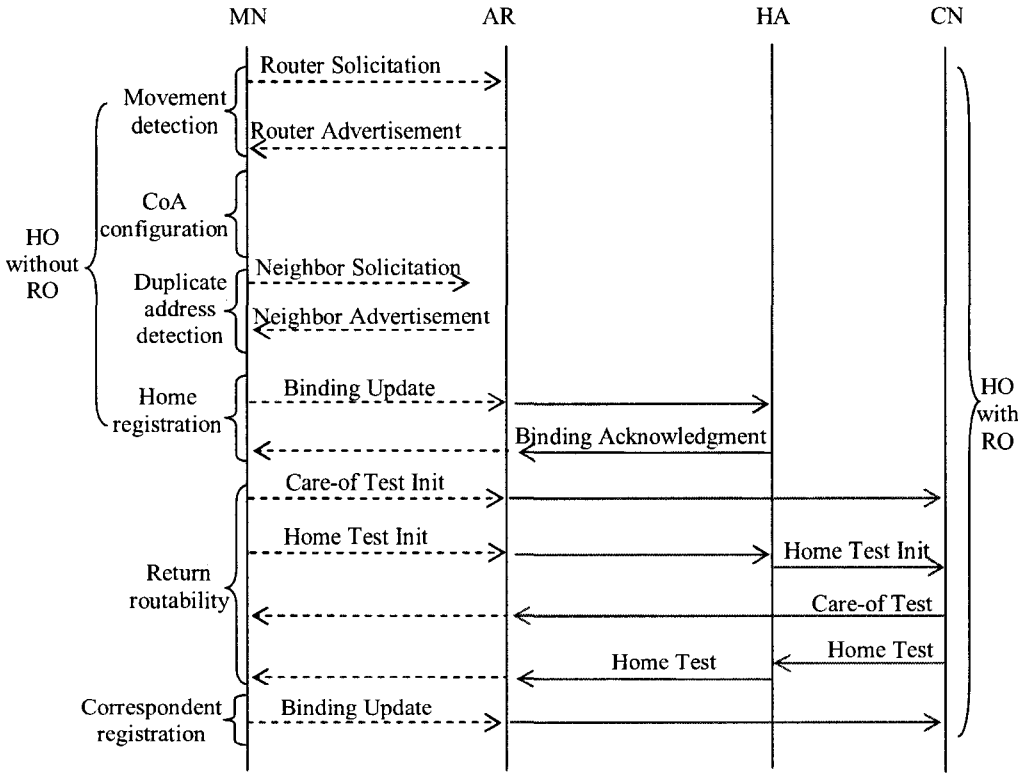

Figure 4.1 Mobility Management Process for MIPv6

In a nutshell, MIPv6 defines an IP-layer mobility management scheme to provide MNs with continuous Internet access while they move from one domain or AP to another. As mentioned above, the process of changing APs is known as handoff. During handoff, there is a period during which MNs are unable to send or receive packets due to delays resulting from both link switching and IP protocol operations. Eventually, handoff latency results in packet loss and degrades network performance, which is unacceptable and detrimental for real-time traffic with user-perceptible service deterioration [59]. As reducing handoff latency could be beneficial to real-time applications, non-real-time applications and throughput-sensitive applications as well, new

enhancements are necessary for optimal performance, especially in situations where mobile users perform local movements and change APs frequently. Moreover, in order to keep bindings updated, an MN periodically informs the HA and CNs about its CoA if and when its location changes or the binding lifetime expires. Under such circumstances, the network is burdened with a huge amount of signalling overhead. Hence, minimizing signalling overhead is also important to improve the network bandwidth usage and extend the battery lifespan of MNs.

4.2.2 Hierarchical mobile IPv6 (HMIPv6)

Even though MIPv6 provides appropriate macro-mobility management features, certain pitfalls pertaining to mobility management procedures remain, such as high signalling overhead, long handoff latency and unacceptable packet loss rate. Hence, improvements are required to enhance the performance of MIPv6. In this context, hierarchical mobile IPv6 (HMIPv6) was designed to reduce signalling overhead and location update delays in MIPv6 using hierarchical mobility agents called mobility anchor points (MAPs). Such entities control the local movements of MNs [8] [9].

When an MN enters a new MAP domain, it receives RAs from access routers. Based on the received information, the MN configures two IP addresses: a regional CoA (RCoA) on a specific MAP subnet and an on-link local CoA (LCoA). After performing a DAD for the LCoA, the MN sends a *local BU* (LBU) message to the MAP in order to establish binding between the RCoA and the LCoA. Upon receiving this message, the MAP performs a DAD for the MN's RCoA and returns a *BAck* message to the MN.

Once the MN moves within the local MAP domain, it only needs to inform the MAP of its new LCoA without further updating the binding at the HA and all CNs. However, if the MN performs inter-domain movements, i.e. moving from one MAP domain to another, the MN first needs to register new binding at the new MAP. After receiving the BAck from the new MAP, the MN sends a *BU* to the HA to establish binding between its home address and the new RCoA. With route optimization, the MN also needs to carry out return routability and correspondent registration procedures in order to create (or update) the identical binding for every CN. Packets addressed to the MN are intercepted by the MAP and tunnelled to the MN's current LCoA. Figure 4.2 illustrates the mobility management procedure for HMIPv6.

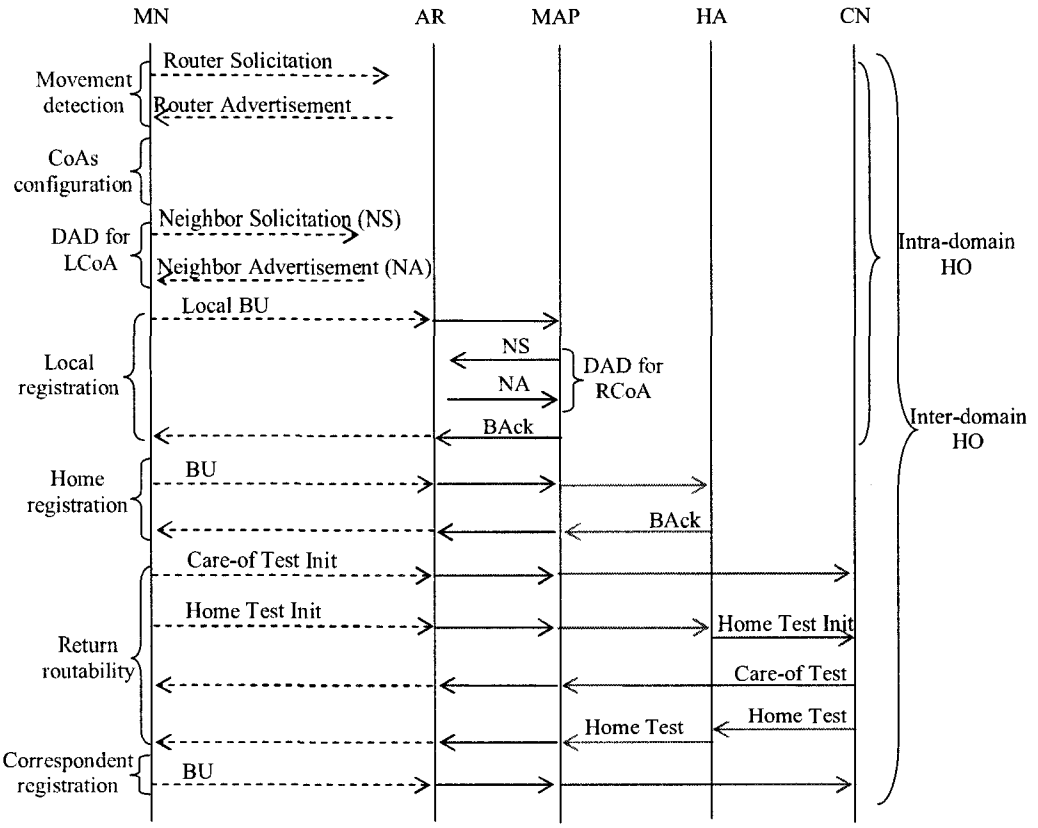

Figure 4.2 Mobility Management Process for HMIPv6

On the whole, the HMIPv6 protocol confines BUs to a local MAP for intra-domain movements, rather than propagating back to the HA and all CNs, thus improving the MIPv6 performance. However, this approach still needs further improvements to support real-time applications as HMIPv6 is mainly concerned with BU delays pertaining to intra-domain roaming, which does not improve the latencies caused by movement detection, CoA configuration as well as CoA verifications [19].

4.2.3 Fast handovers for mobile IPv6 (FMIPv6)

Fast handovers for mobile IPv6 (FMIPv6) enables MNs to rapidly detect their movements and formulate a prospective CoA while still connected to their current subnets [9] [10]. It also provides MNs with the opportunity to use available link-layer event notifications (or triggers) to accelerate network layer handoffs. Consequently, delays

due to network prefix discovery and the generation of new IP addresses are completely eliminated during handoff. Moreover, a bidirectional tunnel is setup between the previous access router (PAR) and new access router (NAR) to avoid packet drops during handoff. In addition, the PAR maintains binding between the MN's previous CoA (PCoA) and the new CoA (NCoA). Hence, packets addressed to an MN are intercepted by the PAR and tunnelled to the MN's NCoA. Such an approach avoids BUs for the HA and all CNs in the event of route optimization.

FMIPv6 supports two operation modes: the predictive and the reactive modes. The former implies that MNs receive fast binding acknowledgement (FBAck) messages from their previous links. As for the latter, MNs do not receive FBAck messages from their attached PARs.

After discovering one or more nearby APs, an MN sends a *router solicitation for proxy advertisement* (RtSolPr) message to the PAR (the MN's default router prior to handoff) to resolve AP identifiers to subnet router information. In response, the PAR sends a *proxy router advertisement* (PrRtAdv) to the MN to indicate whether or not it holds information regarding the NAR. Upon receiving the PrRtAdv, the MN formulates a prospective CoA and sends a *fast binding update* (FBU) to the PAR.

The PAR sends the NAR a *handover initiate* (HI) message to set up a bidirectional tunnel. In return, the NAR validates the proposed NCoA via DAD. If the NCoA is invalid, the NAR allocates an NCoA for the MN and sends the PAR a *handover acknowledge* (HAck). The PAR then binds the MN's PCoA to the NCoA, and forwards an *FBAck* message to the MN. Afterwards, the PAR intercepts packets destined for the MN's PCoA and tunnels them to the NCoA. The NAR then buffers these packets. Upon receiving the FBAck, the MN disconnects from the PAR and initiates link-layer switching procedures. Once attached to the new link, the MN sends an *unsolicited neighbor advertisement* (UNA) to the NAR [11], so that both arriving and buffered packets can be forwarded immediately to the MN. Figure 4.3 shows the predictive mobility management procedure for FMIPv6.

Predictive mobility management enables MNs to receive FBAck on their previous link. Furthermore, receiving FBAck from the PAR means that packet tunneling is already in progress when MNs handover to the NAR. If MNs do not receive the FBAck message on the previous link, they must follow the reactive mobility management procedures.

Performing reactive mobility management can occur for two reasons: either the

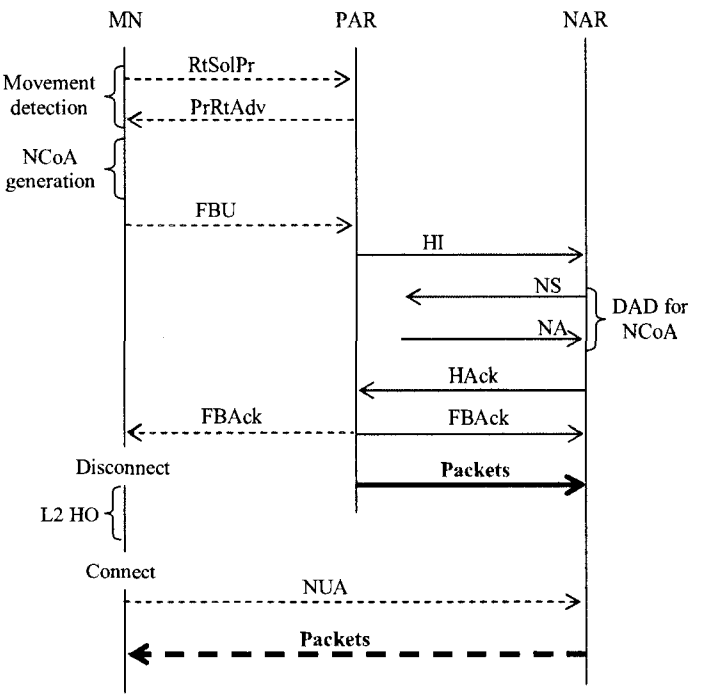

Figure 4.3 Mobility Management Process for Predictive FMIPv6

MN did not send the FBU, or the MN left the link after sending the FBU, which was possibly lost, but before receiving the FBAck from the PAR. Without receiving an FBAck, the MN cannot be sure that the PAR successfully processed the FBU or that the tunnels are ready to use when it handovers to the NAR. As a consequence, the MN sends a *UNA* to the NAR immediately after attaching to the new link and launches a DAD probe for its NCoA. Upon receiving the UNA, the NAR can detect that the designated NCoA is in use when processing the UNA. In this case, it sends the MN a *router advertisement* (RA) with a neighbor advertisement acknowledge (NAACK) option where the NAR assigns an alternate IP address for the MN to use.

Once the MN confirms the validity of its NCoA either via DAD or provided by the NAR, it sends a *neighbor advertisement* (NA) to the all nodes multicast address to join the multicast group. The MN then sends an FBU to the PAR via the NAR. Upon receiving the FBU, the PAR binds the MN's PCoA to NCoA. If necessary, HI and HAck messages are exchanged to establish a bidirectional tunnel between the PAR and NAR. Accordingly, packets destined to the MN's PCoA are intercepted by the PAR and tunnelled to the NCoA via the NAR. Moreover, the *FBAck* message can

also be piggybacked with these packets. Figure 4.4 illustrates the reactive mobility management procedures.

Basically, FMIPv6 addresses the following issues: how can mobile users send packets as soon as they detect a new subnet link? How can they receive packets as soon as the NAR detects their attachment? As the FMIPv6 protocol utilizes pre-handover triggers, its performance, in terms of number of lost packets, depends dramatically on the pre-handoff trigger time, thus becoming unreliable when the pre-handoff trigger is delivered too closely to the actual link switch [59].

4.2.4 Fast handover for hierarchical mobile IPv6 (F-HMIPv6)

In their own ways, both FMIPv6 and HMIPv6 are designed to improve MIPv6 performance in terms of handover delay and signalling overhead, thus making it necessary to amalgamate these two schemes. However, simply superposing FMIPv6 over HMIPv6 induces unnecessary processing overhead for PAR re-tunnelling, as well as inefficient use of network bandwidth [16]-[19]. Hence, an effective integration, called fast handover for hierarchical MIPv6 (F-HMIPv6), is designed to enable an MN to exchange handoff related signalling messages with a local MAP, and a bidirectional tunnel is established between the MAP and NAR, instead of between the PAR and NAR.

Upon receiving layer two (L2) handoff anticipation or triggers, the MN sends a *router solicitation for proxy advertisement* (RtSolPr) message to the selected MAP [17]. This message includes information about potential NAR's MAC address or its identifier. It is assumed that the MAP already knows ARs' network prefixes and MAC addresses within its domain. Hence, in return to the RtSolPr, the MAP sends a *proxy router advertisement* (PrRtAdv) message to the MN. Such message contains either NAR's network prefix or a new on-link local CoA (NLCoA). The former allows the MN to auto-configure an IPv6 address for its interface via a stateless way [43] while the latter is used for stateful address configuration [46]. Subsequently, the MN sends a *fast binding update* (FBU) message to the MAP indicating its previous local CoA (PLCoA) and the NAR's IP address.

After receiving the FBU message from the MN, the MAP sends a *handover initiate* (HI) message to the NAR to establish a bidirectional tunnel. In response to the HI, the NAR sets up a host route entry for the MN's PLCoA and then responds with a *handover acknowledge* (HAck) message. As a result, a bidirectional tunnel is

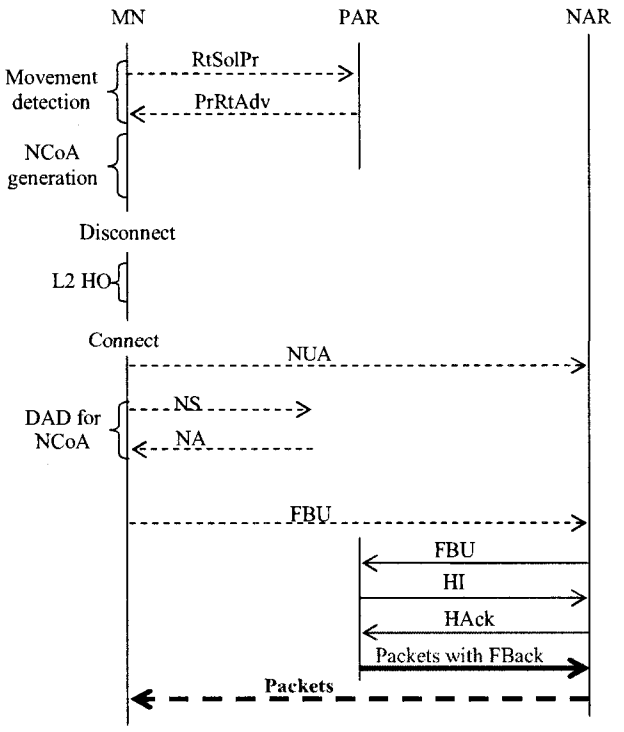

Figure 4.4 Mobility Management Process for Reactive FMIPv6

established between the MAP and NAR.

The MAP sends a *fast binding acknowledgment* (FBAck) message toward the MN over its PLCoA and the NLCoA. Then, the MAP begins to forward the packets destined to the MN to the NAR by using the established tunnel. Once attached on the new link, the MN sends a *fast neighbor advertisement* (FNA) message to the NAR, which then delivers the buffered packets to the MN.

Afterwards, the MN follows normal HMIPv6 operations by sending a *local binding update* (LBU) to the MAP via the NAR. When the MAP receives the new LBU with an NLCoA from the MN, it stops forwarding packets to the NAR and removes the established tunnel for fast handover. In response, the MAP sends a *local binding acknowledgment* (LBACK) to the MN. As a result, the remaining data path follows the HMIPv6 procedures [8] [9]. Figure 4.5 depicts the mobility management procedure in F-HMIPv6.

Even though F-HMIPv6 allows mobile users to benefit from both FMIPv6 and HMIPv6, the handoff latency for intra-domain roaming lasts about $90ms$ whereas

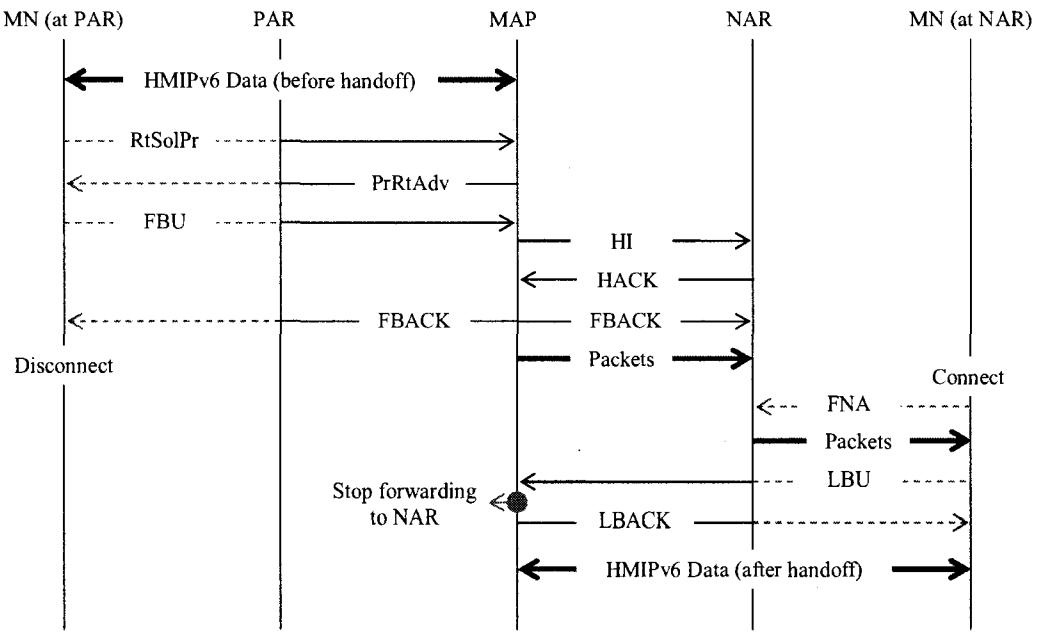

Figure 4.5 Mobility Management Process in F-HMIPv6

the handoff delay for inter-domain roaming rises to about $240ms$ [19], making this protocol unsuitable for multimedia streaming traffic.

Note that all of the above-mentioned mobility management protocols prevent knowing the exact moment when mobile users move or attach to the NARs. Furthermore, as packets are transmitted too early or too late in many cases, this dilemma results in tremendous quantities of lost packets. Simultaneous binding could be an interesting solution for this challenge [60]. Furthermore, MIPv6, HMIPv6, FMIPv6 and F-HMIPv6 protocols consist of host-centric solutions for IP mobility, generating problems when MNs cannot signal their mobility information to the HAs and their correspondents.

4.2.5 Handoff Protocol for Integrated Networks (HPIN)

Based on localized mobility management, access network discovery and fast handoff concepts, a handoff protocol for integrated networks (HPIN) is proposed recently and aims to provide efficient handover management for mobile hosts roaming in next-generation wireless networks [168].

Before the handoff decision is made by an MN, anticipated triggers are generated,

such as high bit error rate, link going down, weak signal strength, etc. After creating these triggers, the MN sends a *candidate access router discovery request* (CARD Req) message to the associated MAP. This message contains user preferences and application-based QoS requirements. The MAP then forwards this message to a network entity, called interworking decision engine (IDE), to request the information about neighboring networks. In the meantime, the MAP obtains detailed information about candidate ARs by exchanging *router information exchange request* (RIX Req) and *RIX reply* (RIX Rep) messages between them. The MAP also sends a *user profile request* to the IDE, which then replies with a *user profile reply* message. After obtaining the user's profile from the IDE, the MAP performs pre-filtering operations, which is based on the user's preference, the required QoS parameters, etc. Then the MAP sends a *CARD reply* (CARD Rep) message to the MN indicating a list of candidate ARs.

When the MN receives a L2 trigger, it selects a target AR from the list, and sends a *FBU* to the MAP. Upon receiving of this message, the MAP sends an *HI* to the NAR, which then performs the DAD procedure and replies the MAP with a *HAck*. Then the MAP forwards an *FBAck* to the MN via its PCoA and NCoA. At the same time, the MAP binds the MN's PCoA to NCoA, and tunnels any packets addressed to the PCoA to the NCoA in the NAR's network.

The MN sends a *router solicitation with FNA* after attached to the new link. In response, the NAR forwards the buffered packets to the MN, and sends an *LBU* to the MAP, which then replies with an *LBAck*. The NAR then forwards this message to the MN, thus completes the handoff process. Figure 4.6 illustrates the intra-domain handoff process for HPIN.

In case of inter-domain movements, bidirectional tunnels are created between the previous MAP (pMAP) and the NAR via the new MAP (nMAP). After receiving the FBU from the MN, the pMAP sends a *handoff request (HOReq) with HI* to the nMAP, which then obtains the user's profile from the IDE through exchanging a pair of messages (*user profile request* and *user profile reply*) between them. The nMAP also sends a *context transfer request* (CTReq) to the IDE for MN-related session management parameters. The IDE then replies with a *context transfer data* (CTD) to the nMAP. Note that the nMAP forwards the *HI* to the NAR, which then replies with an *HACK*. Upon receiving the HACK, the nMAP sends a *Handoff Response* (HORep) to the pMAP. The pMAP sends an *FBAck* to the MN via its PCoA and

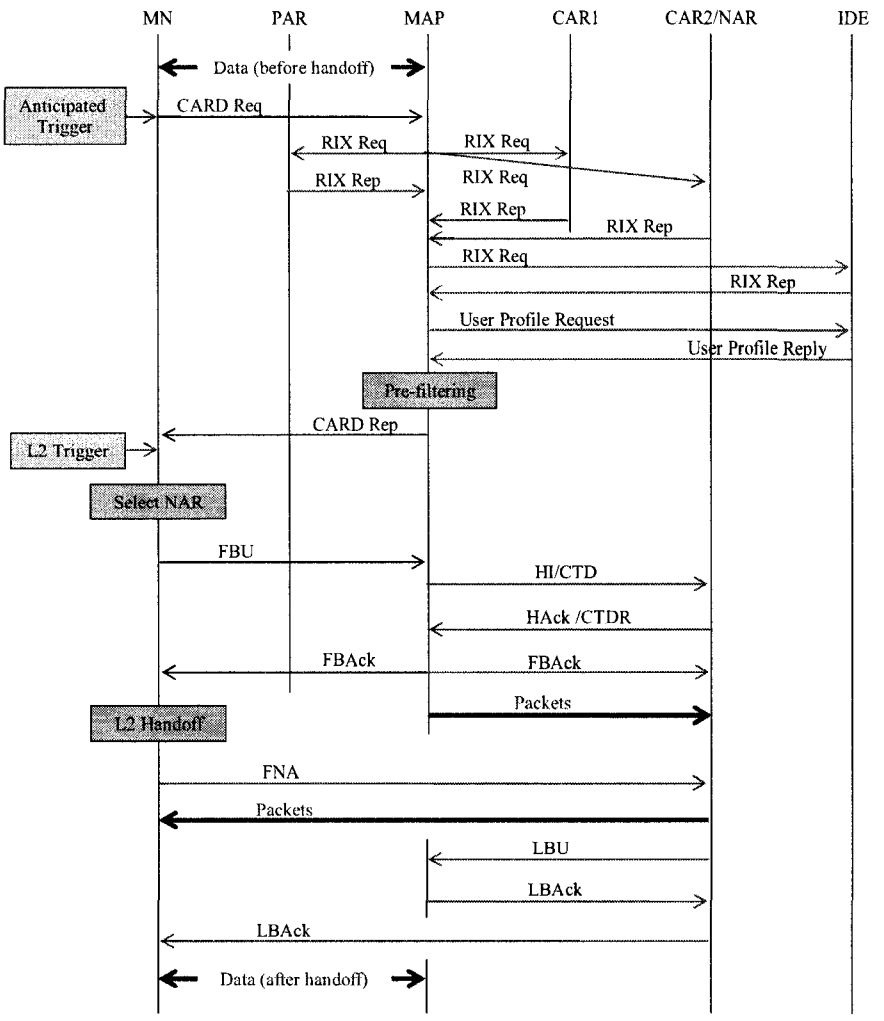

Figure 4.6 Intra-domain Mobility Management for HPIN

the NCoA, and tunnels packets destined to the MN to the NAR. The NAR buffers these packets. When the MN connects to the new link, it sends an *FNA* to the NAR, which then forwards the packets to the MN. The NAR then sends an *LBU* to the nMAP, which then responses with an *LBAck*. The MN then follows the normal MIPv6 registration procedure. Figure 4.7 shows the inter-domain handoff process for HPIN.

HPIN introduces a lot of signalling messages between the MAP and candidate ARs, thus resulting in higher signalling overhead, compared with F-HMIPv6. In addition, this protocol is based on the assumption that the IDE contains user profile

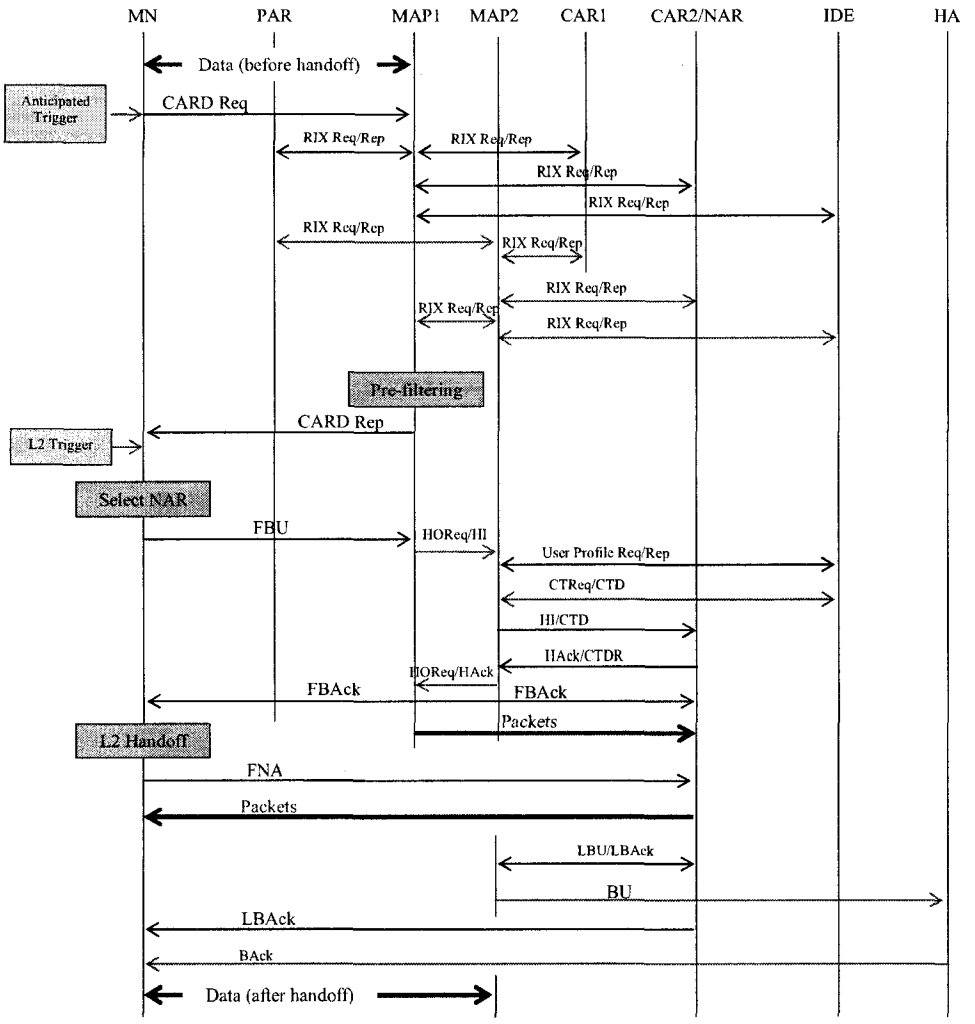

Figure 4.7 Inter-domain Mobility Management for HPIN

for each MN from different service provider. Hence, leads to the problem of scalability. Since network operators are reticent on revealing their clients' database to other operators, this assumption makes the HPIN protocol impractical. And the obtained numerical results are questionable as more signalling messages are required during handoff, compared to F-HMIPv6.

4.2.6 Proxy mobile IPv6 (PMIPv6)

MIPv6 requires certain IPv6 client functionalities for MNs, i.e. MNs must be capable to manage mobility by themselves. However, as certain MNs lack such functionalities,

proxy mobile IPv6 (PMIPv6) is designed to enable network-based mobility management for MNs without their participation in mobility related signalling activities [20].

PMIPv6 extends MIPv6 signalling messages and reuses the functionality of HA to support mobility for MNs without host involvement. In the network, mobility entities are introduced to track the movements of MNs, initiate mobility signalling on behalf of MNs and setup the routing state required. Figure 4.8 illustrates the handoff management procedure for PMIPv6.

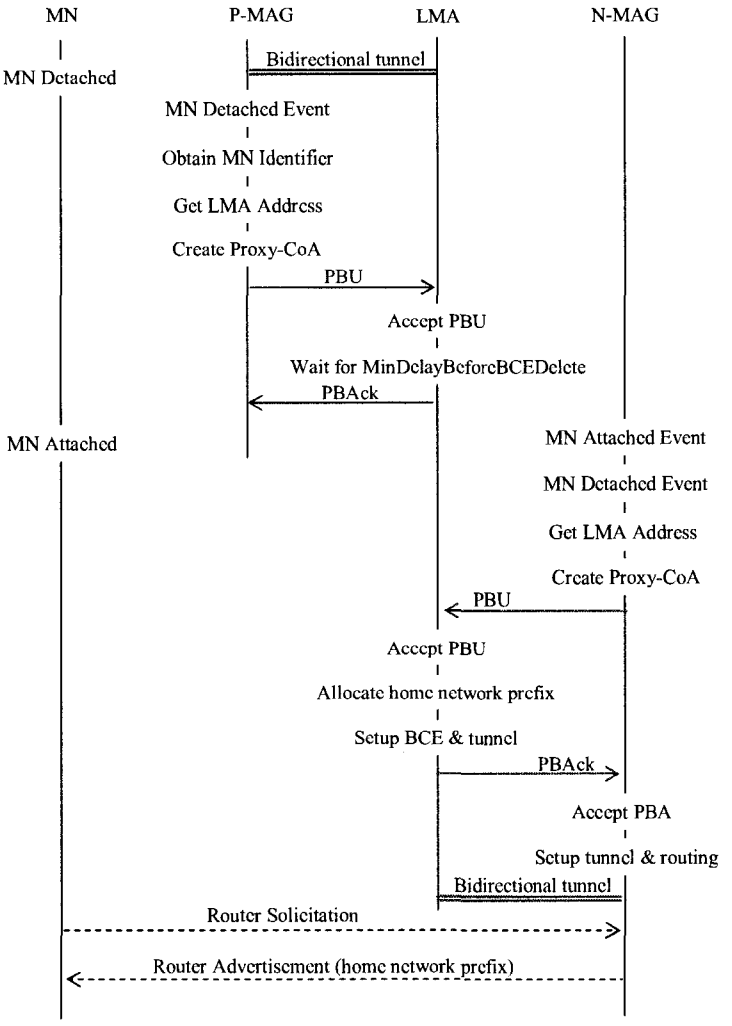

Figure 4.8 Mobility Management Process for PMIPv6

When an MN attaches to a new access link connected to a mobile access gateway (MAG), the MAG authenticates and authorizes the MN before providing PMIPv6

services. While access is authenticated or network attachment events are notified, the MAG acquires the MN's identifier and uses it to access the MN's policy store and retrieves the address of the local mobility anchor (LMA) that serves as the MN's HA.

The MAG also configures a proxy CoA for the MN, and then sends a proxy binding update (PBU) message to the LMA. In return, the LMA updates its binding cache entry (BCE) for that MN and replies with a proxy binding acknowledgment (PBAck) that contains the MN's home network prefix assigned by the LMA. Upon receiving the PBAck, the MAG establishes a bidirectional tunnel between its proxy-CoA and the LMA address. Then, the MAG periodically sends router advertisement (RA) messages to the MN on the access link advertising the MN's home network prefix as the hosted on-link prefix. Or, the MN can send a router solicitation (RS) to request for an immediate RA from the MAG. In return, the MAG replies with a RA equipped with the MN's home network prefix. Hence, the MN can configure a home address using either stateful or stateless mode depending on the modes permitted on that access link. As a result, the MN only uses its home address for all its communication and the proxy-CoA is invisible to the MN. Packets sent to the MN are intercepted by the LMA, tunnelled to the MAG which removes the tunnel header before forwarding them to the MN. Likewise, packets sent by the MN are intercepted by the MAG and tunnelled to the LMA which then removes the outer header and forwards them to the CN.

When an MN handoffs from a previous MAG (P-MAG) to new MAG (N-MAG), the MN detached event is detected by the P-MAG, which sends a PBU to the LMA in order to update the MN's binding and routing states. After receiving the PBU, the LMA waits for a certain $MinDelayBeforeBCEDelete$ amount of time before deleting the MN's entry from its BCE. Then, the LMA sends a PBAck to the P-MAG.

When the MN attaches to the new link, the N-MAG obtains its identifier either via access authentication or notified network event. Then, the N-MAG accesses the MN's profile and obtains the LMA address. The N-MAG also configures a new proxy-CoA for the MN on one of its interfaces. Subsequently, it sends a PBU to the LMA which assigns a home network prefix to the MN and sends a PBAck to the N-MAG that contains this home network prefix. Upon receiving the PBAck, the N-MAG sets up a bidirectional tunnel with the LMA before configuring a data path for the MN's traffic. Finally, it emulates the MN's home link by sending RA messages to the MN

that contains the MN's home network prefix.

As PMIPv6 was only recently drafted by the IETF community, certain shortcomings remain to be addressed, such as triangular routing problems without route optimization support, security concerns, etc. Furthermore, since detecting MNs' detachment and attachment events remains difficult in many wireless networks, in-fly packets will certainly be dropped at N-MAGs. Aside from that, a link-layer access authentication protocol is required to enable the MAG to obtain MNs' identifiers. However, such proposals have not yet been developed.

4.2.7 Summary

All of the aforementioned IP layer mobility management protocols prevent knowing the exact moment when mobile users detach from the PAR and when they attach to the NAR. However, in most cases, packets are transmitted either too early or too late between the PAR and NAR, this dilemma results in tremendous quantities of lost packets. Simultaneous binding (bi-cast or n-cast) could be an interesting solution for this challenge [60], [93]. However, such method introduces higher signalling overhead and traffic loads into the network, because packets destined to the MN are forwarded by the PAR to one or more future locations of the MN even before the MN actually moves to that location. Furthermore, using simultaneous binding, ARs need more buffer spaces for the storage of data packets, thus wasting network resources. On the other hand, MIPv6 [7], HMIPv6 [8] [9], FMIPv6 [10] [11] and F-HMIPv6 [16]-[19] protocols consist of host-centric solutions for IP mobility, generating problems when MNs cannot signal their mobility information to the HAs and their correspondents.

4.3 Research direction

Aside from the IETF's standardization activities, one of the well-known activities is that the IMS-based solution for service provisioning and handover between fixed and mobile operators is to be used in NGWNs. Within the 3GPP [61] and the 3GPP2 [62], research is well underway to provide handoff capabilities between disparate technologies and different domains. This 3GPP activity falls under the projects undertaken by system architecture evolution (SAE) and long term evolution (LTE).

The IEEE 802.21 working group [63] delves into the media independent handover

(MIH) project while IEEE 802.11u [64] addresses inter-technology handovers. The IEEE 802.11r [65] task group is currently attempting to enhance handoff performance within IEEE 802.11. Additionally, a number of research and development activities are ongoing by academia, research institutes and industries simultaneously. To date, numerous research projects have focused on certain facets pertaining to mobility management, such as security, QoS provisioning and routing. Nevertheless, it is essential to develop an end-to-end mobility management solution that covers the entire span of the network, from radio access to packet delivery with guaranteed QoS. In the near future, mobile users carrying an integrated terminal can use a wide range of applications provided by multiple wireless networks [2].

Fourth generation (4G) wireless networks are expected to integrate a large number of heterogeneous wireless technologies towards universal seamless access and omnipresent computing through seamless mobility that enables users to access sought services anywhere, at any time [66]. One of the main research challenges for seamless mobility consists of providing available and reliable intra- and inter-system handoff solutions [67]. This implies the design of new and efficient mobility management schemes to optimize QoS while providing flawless mobility.

Moreover, as 4G wireless networks tend to revolve around universal mobile access (UMA) and ubiquitous computing via seamless mobility, another major challenge for seamless mobility lies in the design of an effective and efficient vertical handoff protocol for MNs that move between different types of networks. As traditional operations for handoff detection policies, decision metrics and radio link transfers cannot adapt to dynamic handoff criteria, nor deliver context-aware services or ensure network interoperability, it is obvious that new techniques are required to manage inter-system mobility [68].

This section presents some basic concepts and background related to IP layer mobility management protocols proposed by the working groups of the IETF. MIPv6 [7], the baseline protocol, made significant contributions to enable MNs access to the Internet while roaming around. However, the MIPv6 mobility management procedure leads to impaired handoff performance in terms of signalling load, handoff delays and packet losses. Hence, improvements are crucial to enhance network performance. In this context, HMIPv6 [8] [9], FMIPv6 [10] [11], F-HMIPv6 [16]-[19] and PMIPv6 [20] were designed for localized mobility management. All of these improvements have their own merits and shortcomings. Furthermore, several standard-

ization activities are underway to create an intelligent mobility management scheme that enables mobile users to benefit from its cost-effectiveness, enhanced features, location-independence and seamless mobility across different systems via multi-mode radio interfaces.

4.4 Proposed seamless schemes for IP layer handoff management

Based on the integrated architecture, we propose seamless handoff schemes for mobile users in next-generation heterogeneous wireless networks. The basic idea is to pre-configure bidirectional secure tunnels (BSTs) among adjacent ARs before actual handover. The QoS related parameters (such as delay jitter, packet loss) and security aspects (like authentication method, tunnelling keys) are specified for each unidirectional tunnel in a context of contracted service level agreement (SLA), which is made among telecommunication operators. These SLAs enable service operators to provide services to mobile users from other operators on condition that the involved two parties have made an SLA within which the mobile users are given the opportunity to exploit pre-configured BSTs for their ongoing multimedia sessions during handoff. In addition, with the assistance of pre-established bidirectional tunnels, mobile users can utilize their previous valid IP addresses in a new visiting network or domain. Hence, minimizing the service interruption during handoff. Additionally, new routing policies are added to ARs, which enables the delivery of packets to mobile users that have a topologically invalid address within the attached network.

The proposed seamless handoff schemes for mobile IPv6 (SMIPv6) comprise two stages: making bilateral SLAs between neighboring radio access networks (RANs) prior to actual handoff and using such SLAs during handoff.

The first stage consists of making SLAs among adjacent ARs. Such SLAs enable RANs to establish business and security relationship with their neighborhood. Consequently, communication services are offered to a list of subscribers from other mobile operators. Within an SLA, the conditions of bidirectional tunnel are specified according to the traffic type. For example, QoS related parameters, traffic classification aspects, QoS mapping algorithm and shared keying materials are elaborated in detail. As a result, bidirectional tunnels are established prior to actual handovers.

So signaling for setup such tunnels is eliminated completely during fast handovers.

The second stage consists of utilizing the pre-configured bidirectional secure tunnels during handoff. In doing so, new routing policies are added to the ARs. During handoff, the NAR may receive a type of data packets from the PAR, shown in Table 4.1.

Table 4.1 An Example of a Packet Sent by the PAR and Received by the NAR

Source Address 1	Destination Address 1	Source Address 2	Destination Address 2	Token	Data
PAR's IP address	NAR's IP address	CN's IP address	MN's PCoA	token 1	data

Upon receiving such a packet, the NAR usually decapsulates it and verifies its ultimate destination: MN's previous care-of address (PCoA). According to the normal IP routing policy, the NAR forwards the packet to the PAR, because the destined node MN is supposed to be located in the PAR's subnet. This leads to loop routing problem presented in FMIPv6. To resolve the problem, our proposed seamless handoff schemes force the NAR to buffer such packets and wait for the attachment of the MN.

During handoff, the NAR may receive another type of data packets from an MN, shown in Table 4.2. Generally, such kind of packets are dropped by the NAR due to ingress filtering. However, using SMIPv6, the NAR encapsulates such packets and forwards them to the PAR, which then decapsulates and sends them to the CN after successful decryption and authentication.

Table 4.2 An Example of a Packet Sent by an MN and Received by the NAR

Source Address	Destination Address	Encrypted Data
MN's PCoA	CN's IP address	encrypted data using pre-shared key between MN and PAR

Since SMIPv6 empowers MNs to use their valid PCoAs, MNs related context information can be kept intact at the PAR. Hence, delays caused by the context transfer process is removed completely during handoff. In Addition, the MN can resume or initiate its communication on the new link using its valid PCoA with the help of pre-configured tunnels within the SLA.

Compared with the bidirectional edge tunnel handover for IPv6 (BETH) [91], SMIPv6 does not require *handover request* and *handover reply* messages exchanged to set up bidirectional tunnels during handoff; neither does it exploit link-layer pre-triggers to facilitate IP layer handoff. Given that both FMIPv6 and BETH protocols utilize pre-handover triggers, their performance, in terms of the number of lost packets

and handoff latency, depends greatly on the pre-handoff trigger time, thus becoming unreliable when the pre-handoff trigger is delivered too closely to the actual link switch [59], [92].

4.4.1 Proposed seamless handoff schemes (SMIPv6)

We assume that an MN roams from an AR to another in the IPv6-based wireless networks. And it acquires a valid care-of address (CoA) within the range of the first AR, called PAR. And such CoA is named previous CoA (PCoA). In addition, the MN establishes a security association (SA) with the PAR, and configures a shared secret key.

In the overlapping zone of the PAR and its neighboring ARs, the MN receives one or more beacons from nearby APs. Such beacons contain the AP's identifier (AP-ID). In case of horizontal handoff (handoff within the same access technology), the MN may select the most suitable AP according to the received signal strength. In case of vertical handoff (handoff among different access technologies), the MN may select the most appropriate AP using the score function presented in [68]. Thereafter, the MN sends a *seamless binding update* (SBU) message to the PAR before breaking its connection between them. Such message contains the selected new AP's identifier (NAP-ID) and a session token. The token is generated by the MN and used to avoid replay attack.

We also assume that the PAR has preknowledge about its neighboring ARs, such as their associated APs' identifier, the AR's IP address, etc. Therefore, upon receiving the *SBU*, the PAR maps the NAP-ID to the associated AR's IP address and starts intercepting packets addressed to the MN. The PAR caches one copy of these packets, at the same time, it begins tunnelling them to the NAR. The session token is then inserted into the first tunnelled packet. In a meanwhile, the PAR adds an entry to its *Forwarding Tunnels List.* Such list is utilized to track the state of tunnels.

Note that packets buffered by the PAR will be forwarded to the MN in case of *ping pong* and *erroneous* movements. The former implies that MNs move between the same two ARs rapidly while the latter connotes that MNs think that they enter into a NAR network, but they are actually either moving to a different AR or abort their movements by returning to the PAR.

Upon receipt of the tunnelled packets from the PAR, the NAR decapsulates and

buffers them. At the same time, the NAR adds an entry into its *Forwarding Tunnels List*. The session token is extracted from the first tunnelled packet and added by the NAR into its *Token List*, which is indexed by MN' IP address. At the meantime, the NAR creates a host route entry for the MN's PCoA, and allocates a unique NCoA for the MN. Here we advocate that each AR manages a private address pool and guarantees the uniqueness of each individual address from the pool. By this means, the DAD process is removed from the overall handoff process, thus expediting handoff process and reducing handover latency.

Once attached to the new link, the MN sends a *seamless neighbor advertisement* (SNA) message to the NAR immediately. Besides all the fields in the message format of *unsolicited neighbor advertisement* (UNA) defined in FMIPv6 [11], the message SNA contains the same session token sent by the MN to the PAR and the PAR's IP address. The *IP source address* of the SNA is the MN's PCoA, and its *IP destination address* is typically the all-nodes multicast address. In addition, the *source link layer address (LLA)* is the MN's LLA while the *destination LLA* is the NAP's LLA. Here we assume that the NAP's LLA equals to the NAP's ID, which the MN acquires at the moment of preparing the handoff.

The NAR then checks its *Forwarding Tunnels List* and the buffer for packets addressed to the MN. With the assistance of the *Token List*, the NAR verifies whether the received session token from the MN is the same as the one from the PAR. In case of equality, the NAR forwards packets and the assigned NCoA to the MN. Otherwise, the NAR extracts the PAR's IP address from the received SNA, and generates a *fast binding update* (FBU) message on behalf of the MN, and sends the *FBU* to the PAR. Such message contains the MN's MAC address and its PCoA. The PAR then verifies the MN's identities and returns a *fast binding acknowledgment* (FBAck) to the NAR. At the same time, the PAR adds an entry into its *Forwarding Tunnels List* and *Reverse Tunnels List*, respectively. Upon receiving the *FBAck*, the NAR forwards packets and the NCoA to the MN. Consequently, the MN becomes reachable on the new link under both CoAs: PCoA and NCoA. Figure 4.9 illustrates the predictive mobility management procedure for SMIPv6.

In case where the MN initiates a new communication session with correspondent nodes (CNs) using its PCoA, it uses the pre-shared key with the PAR to encrypt the data packets. Note that instead of using the pre-shared key, the encapsulating security payload (ESP) header may also be used to provide confidentiality, data origin

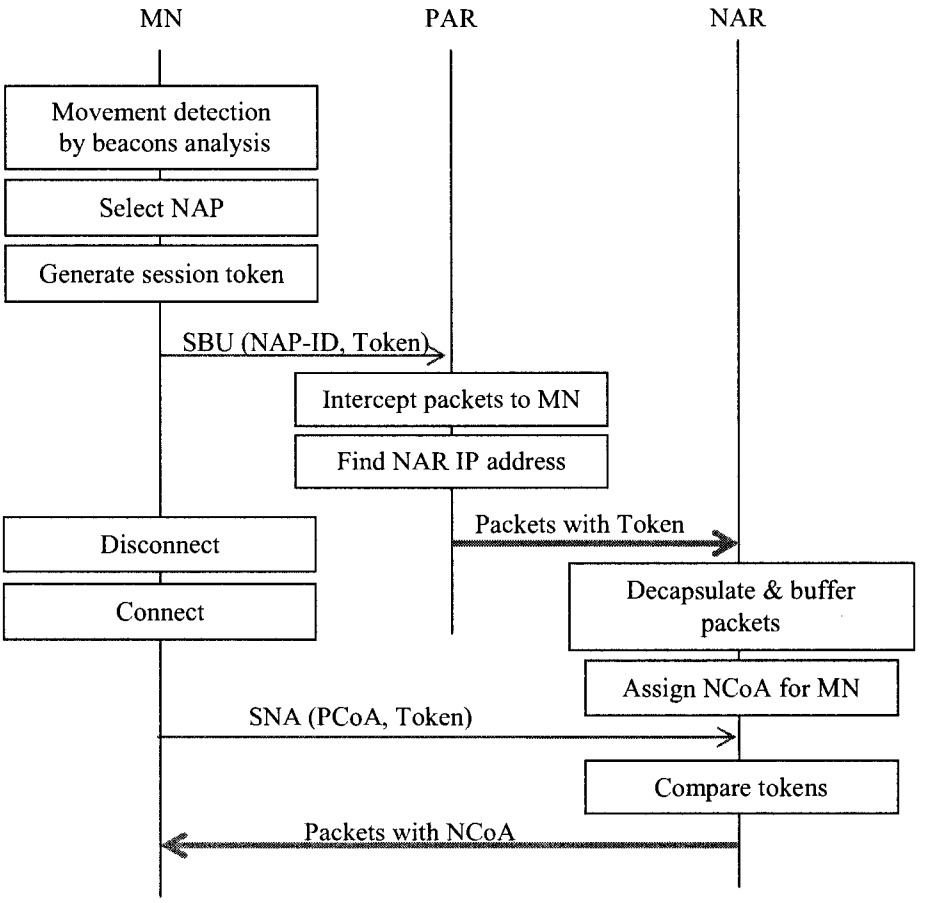

Figure 4.9 Predictive Mobility Management Process for SMIPv6

authentication, connectionless integrity, an anti-replay service, and limited traffic flow confidentiality [48].

And then, these packets are intercepted by the NAR. Before executing ingress filtering, the NAR extracts the subnet prefix information from the PCoA, and checks whether or not there is a pre-established SLA with the PAR using its *contract database* (CD). In case where the MN is on the list of subscribers specified in the SLA, the NAR begins tunnelling the MN's packets to the PAR, and adds an entry into the *Reverse Tunnels List* for further tunnel maintenance and billing issues. Otherwise, the NAR discards the intercepted packets.

Upon receipt of the tunnelled packets from the NAR, the PAR decapsulates the packets, performs ingress filtering for the MN's PCoA. And then, PAR decrypts packets using the pre-shared key with the MN. Afterwards, the PAR forwards the

decrypted packets to the CN and adds an entry into its *Reverse Tunnels List.* Figure 4.10 illustrates the reactive mobility management procedure for SMIPv6.

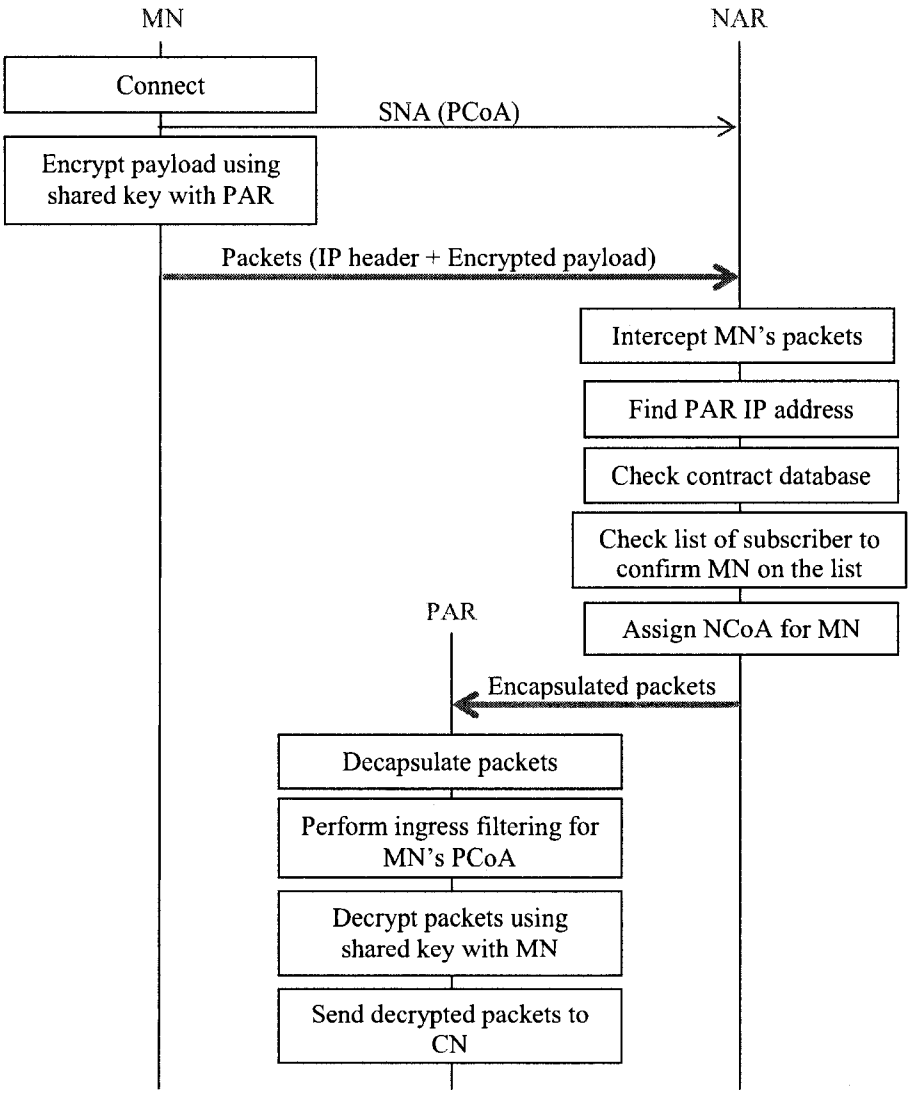

Figure 4.10 Reactive Mobility Management Process for SMIPv6

Note that in FMIPv6, even though the MN is IP-capable on the new link, it cannot use NCoA directly with a CN before the CN first establishes a binding cache entry (BCE) for the MN's NCoA. However, our proposed fast and seamless scheme bypasses this problem by allowing MNs to utilize its valid PCoA immediately after connected to the new link. Hence, SMIPv6 provides not only expedited forwarding packets to

MNs but also accelerated sending packets to their correspondents, thus optimizing handoff performance in terms of signalling overhead, handoff latency, packet drop rate, etc. In addition, SMIPV6 is independent of the architecture. For example, if bidirectional tunnels are pre-configured between adjacent MAPs, SMIPv6 is also applicable to reduce handoff delay and packet losses.

Once terminating its ongoing session using the PCoA on the NAR's link, the MN can follow the legacy MIPv6 or HMIPv6 registration procedures. Let T_s denote the average session duration and T_c be the average cell residence time, we assume that $0 \leq T_s \leq 2T_c$. Under this condition, a session will be terminated either within an AR or between two adjacent ARs. So there is no extra overhead at the intermediate ARs in relaying packets and maintaining the Forwarding Tunnel Table.

To facilitate tunnel maintenance, the MN sends a *Tunnel Bye* message to the NAR, which then releases the reserved bandwidth for the MN, and forwards the same message to the PAR. As a consequence, entries in *Forwarding Tunnels List* and *Reverse Tunnels List* are removed or refreshed. However, SMIPv6 requires bidirectional tunnel remains active until MNs complete their binding update procedures with its correspondents, similar idea as that of the FMIPv6 protocol.

Typically, a session is identified by a group of information such as session ID, source address, destination address, source port number, destination port number, etc. When moving from one network to another, an MN loses its network connectivity and becomes unreachable because its previous source address is invalid in the visiting network. Under the circumstances, the MN has to acquire a new CoA and registers the new CoA with its HA and all active CNs. Prior to successful registration, the MN cannot receive and send packets in the foreign network, thus the ongoing session is disrupted during handoff. In case where the MN executes multimedia applications such as video-streaming, it cannot tolerate the degraded quality of the session. SMIPv6 resolves such problem by allowing MNs to utilize their previous IP addresses on the new link without experiencing unacceptable QoS, thus guarantee seamless roaming with ongoing sessions.

4.5 Analytical modeling 1

In this section, we investigate the performance of the proposed seamless handoff schemes using analytical models. To evaluate the handoff performance in the pro-

posed seamless handoff schemes for mobile IPv6 (SMIPv6), we adapt the analytical model in [167] and make the same assumption using the same reference values to get comparable numeric results among MIPv6, FMIPv6, F-HMIPv6 and SMIPv6. The following performance metrics are calculated in this section.

- signalling costs in term of handover latency.
- Packet delivery costs in term of buffer space required during handoff.

4.5.1 signalling cost

Since FMIPv6 [10] is based on anticipation of fast handover using L2 trigger, the signalling costs due to additional messages may vary according to the probability that the handover anticipation is correct [167]. Thereafter, two probabilities are defined as follows:

- P_s: the probability that L3 handover occurs after the L2 trigger
- P_f: the probability that L3 handover doesn't take place after L2 trigger

To simplify the analysis, we use the following denotation:

- S_s: the signalling cost for a successfully anticipated L3 handover.
- S_f: the signalling cost with a failed L3 handover.
- C^t_{MN-PAR}: the transmission cost on a wireless link between the MN and PAR.
- $C^t_{PAR-NAR}$: the transmission cost on a wired link between the PAR and NAR. Since the transmission cost on the wired link is proportional to the hop distance between the involved network entities. To simplify analysis, we define $C^t_{PAR-NAR} = \tau \times d_{PAR-NAR}$ where τ denotes the unit transmission cost on a wired link, $d_{PAR-NAR}$ represents the hop distance between the PAR and the NAR.
- P_z: the processing costs at the network entity z.

In case where no actual handoff happens after the L2 trigger, the messages of RtSolPr, PrRtAdv, FBU, HI and HACK become useless. We assume that $C^t_{MN-PAR} = C^t_{MN-NAR} = \kappa$ and $P_{PAR} = P_{NAR} = P_{AR}$. Let τ denote the unit transmission cost on a wired link. Hence, the signalling cost functions for FMIPv6 in case of a failed L3 handover and in case of successful L3 handoff are formulated as follows:

$$S_{FMIPv6-f} = 3\kappa + 2\tau \times d_{PAR-NAR} + 3P_{AR} \tag{4.1}$$

$$S_{FMIPv6-s} = 5\kappa + 3\tau \times d_{PAR-NAR} + 5P_{AR} \tag{4.2}$$

Let α and β denote weighting factors, the signalling cost function for FMIPv6 is given as follows:

$$C^s_{FMIPv6} = \alpha \times P_s \times S_{FMIPv6-s} + \beta \times P_f \times S_{FMIPv6-f} \tag{4.3}$$

where $P_f = 1 - P_s$

In the same vein, we observe that if there is no real handover after the L2 trigger (e.g. scanning result of the potential AP's L2 ID in a wireless LAN environment), the SBU message become unnecessary. Additionally the proposed SMIPv6 schemes aim to minimize packet loss and handoff latency by allowing mobile users to exploit their PCoAs for the ongoing sessions during handoff. The signalling cost functions for SMIPv6 are formulated as follows:

$$S_{SMIPv6-f} = \kappa + P_{AR} \tag{4.4}$$

$$S_{SMIPv6-s} = 2\kappa + 2P_{AR} \tag{4.5}$$

$$C^s_{SMIPv6} = \alpha \times P_s \times S_{SMIPv6-s} + \beta \times P_f \times S_{SMIPv6-f} \tag{4.6}$$

where $P_f = 1 - P_s$, α and β are weighting factors.

Assuming that the hop distance between the PAR and MAP equals to the one between the NAP and MAP. That is, $d_{PAR-MAP} = d_{NAR-MAP=d_{AR-MAP}}$. Within the same logic, the signalling cost functions for F-HMIPv6 are expressed as follows:

$$S_{FHMIPv6-f} = 3\kappa + \tau \times 5d_{AR-MAP} + P_{AR} + 2P_{MAP} \tag{4.7}$$

$$S_{FHMIPv6-s} = 7\kappa + \tau \times 9d_{AR-MAP} + P_{AR} + 5P_{MAP} \tag{4.8}$$

$$C^{s}_{FHMIPv6} = \alpha \times P_s \times S_{FHMIPv6-s} + \beta \times P_f \times S_{FHMIPv6-f} \quad (4.9)$$

Assume that only a pair of messages (neighbor solicitation and neighbor advertisement) are exchanged for the DAD process and the time taken for new CoA configuration is negligible. Hence, the signalling cost function for MIPv6 is formulated as follows:

$$S_{MIPv6} = 11\kappa + \tau \times (4d_{AR-HA} + 3d_{AR-CN} + 2d_{HA-CN}) + P_{AR} + 3P_{HA} + 3P_{CN} \quad (4.10)$$

Within the same logic, the signalling cost functions for HMIPv6 (intra-domain) is formulated as follows:

$$S_{HMIPv6} = 6\kappa + 4\tau \times d_{AR-MAP} + P_{AR} + 2P_{MAP} \quad (4.11)$$

4.5.2 Packet delivery cost

In this section, we also adapt the analytical model of [167] to study packet delivery cost and the cost associated with both the forwarding packets and the lost packets are considered in term of the additional buffer space used by the packets. Let $C_{forwarding}$ denote the forwarding Cost, and C_{loss} denote the loss cost, the packets delivery cost function is given as follows [167].

$$C^{p} = \delta \times C_{forwarding} + \gamma \times C_{loss} \quad (4.12)$$

where δ and γ are weighting factors.

As we mentioned above, the forwarding cost is defined as the additional buffer space used during the forwarding time, this cost is proportional to the packet arrival rate (λ) and the forwarding time, the same reason is applicable to the loss cost. So we get the following equations:

$$C_{forwarding} = \lambda \times t_{forwarding} \quad (4.13)$$

$$C_{loss} = \lambda \times t_{loss} \quad (4.14)$$

Packet delivery cost for MIPv6

In order to get the values of $t_{forwarding}$ and t_{loss}, Figure 4.11 illustrates the time diagram of the handoff process for MIPv6 [167].

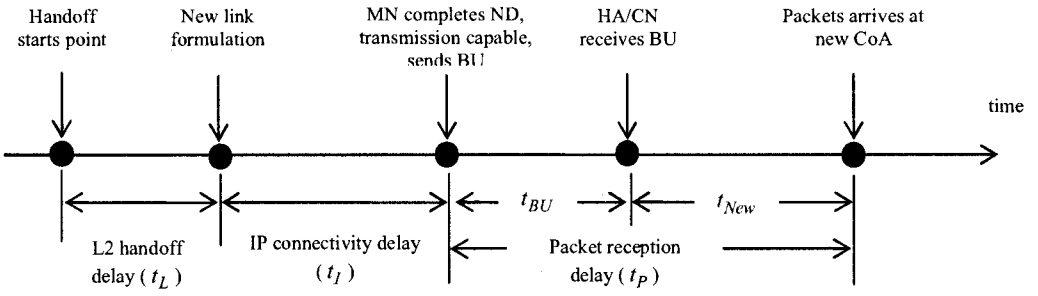

Figure 4.11 Timing Diagram of the Handoff Process in MIPv6

To simplify the analysis, we assume that packet forwarding begins after the neighbor discovery procedure in MIPv6, so the forwarding cost is expressed as follows:

$$C_{forwarding} = \lambda \times t_{forwarding} = \lambda \times (t_{BU} + t_{New}) \tag{4.15}$$

As to the loss cost, because MIPv6 does not support handover anticipation, all packets during $t_L + t_I$ will be lost, so we have:

$$C_{loss} = \lambda \times t_{loss} = \lambda \times (t_L + t_I) \tag{4.16}$$

Hence, the packet delivery cost for MIPv6 can be expressed as follows:

$$C^p_{MIPv6} = \delta \times \lambda \times (t_{BU} + t_{New}) + \gamma \times \lambda \times (t_L + t_I) \tag{4.17}$$

where δ and γ are weighting factors; λ denotes the average packet arrival rate, the values for t_{BU}, t_{New}, t_L and t_I are shown in Figure 4.9.

Packet delivery cost for FMIPv6

In FMIPv6, using the tunnel established after the messages such as RtSolPr/PrRtAdv and HI/HACK, the PAR starts forwarding an MN's packets to the NAR after receiving the FBU message from the MN. The time diagram of the handoff process in

FMIPv6 is shown in Figure 4.12.

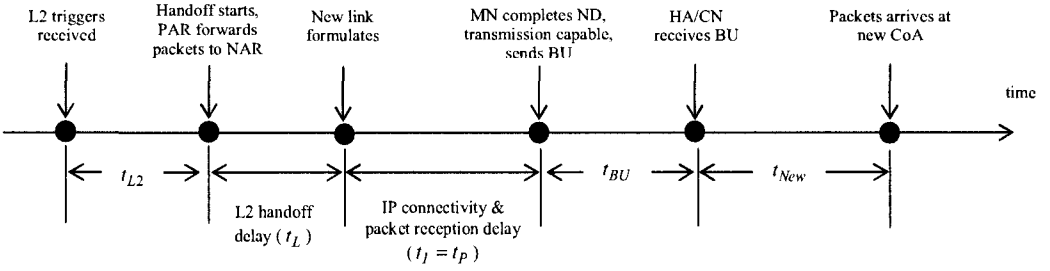

Figure 4.12 Timing Diagram of the Handoff Process in FMIPv6

In this figure, t_{L2} denotes the time taken from the L2 trigger to the starting point of link switching. In order to simplify the performance analysis, we assume that if an MN sends an FBU to the PAR, it will definitely start L3 handover to the NAR without any exception [167]. Under this assumption, all incoming packets after the L2 trigger will not get lost during handover; however, if the MN moves to the new network before the PAR establishing a forwarding tunnel, some packets may also be lost. To define the packet delivery cost function of FMIPv6, we define t_R as the time required to establish the bidirectional tunnel between the PAR and the NAR during handoff. So we have the following equations:

$$C_{forwarding} = \lambda \times t_{forwarding} = \lambda \times (t_L + t_I + t_{BU} + t_{New}) \tag{4.18}$$

$$C_{loss} = \lambda \times t_{loss} = \lambda \times \max\{(t_R - t_{L2}), 0\} \tag{4.19}$$

$$C^p_{FMIPv6} = \delta \times \lambda \times (t_L + t_I + t_{BU} + t_{New}) + \gamma \times \lambda \times \max\{(t_R - t_{L2}), 0\} \tag{4.20}$$

If the hierarchical architecture is deployed and the intra-domain handover is taken into account, we have the following equation:

$$t_{BU} = t_{MN-NAR} + t_{NAR-MAP} + t_{MAP-CN} \tag{4.21}$$

$$t_{New} = t_{CN-MAP} + t_{MAP-NAR} + t_{NAR-MN} \tag{4.22}$$

Note that t_{x-y} denotes the time taken from the network entity x to y. In order to simplify the analysis, we assume that $t_{x-y} = t_{y-x}$ and the transmission time on the same link for different packets is the same, thus $t_{BU} = t_{New} = t_{CN-NAR} + t_{NAR-MN}$.

Packet delivery cost for SMIPv6

With the purpose of getting comparable numerical results, we assume that if an MN sends an SBU message to the PAR, it will definitely start L3 handover to the NAR without any exception [167]. In addition, to avoid packet losses, we assume that both the PAR and NAR have infinite buffer so that the packets could not be dropped because of the buffer overflow. Figure 4.13 illustrates the timing diagram of the handoff process in SMIPv6.

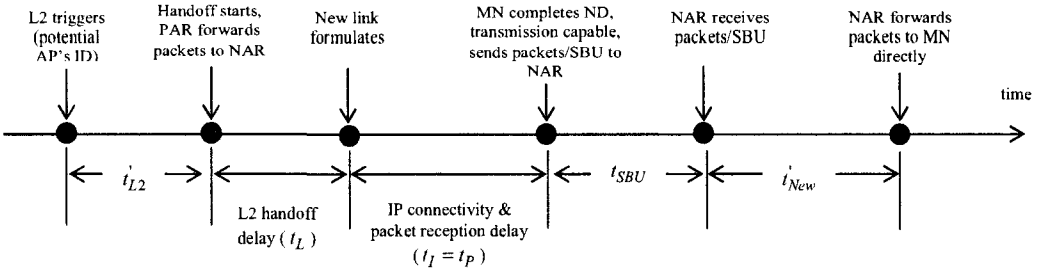

Figure 4.13 Timing Diagram of the Handoff Process in SMIPv6

From Figure 4.13, we observe that in SMIPv6, with the help of access routers tunnelling protocol (ARTP), the bidirectional secure tunnels among adjacent ARs are established before actual handover. Hence, the exchanged messages (HI and HACK) to configure such BSTs become unnecessary during handoff. Moreover, since SMIPv6 schemes allow MNs to utilize their previous CoAs in the new visiting network, the messages (RtSolPr and PrRtAdv) used to help MNs to formulate their new CoAs on the new link are also unnecessary. Furthermore, MNs with ongoing multimedia applications are not required to send BU messages to their HAs and all active CNs during handoff. Thus In SMIPv6, $t_{BU} = 0$. Additionally, the packets from a CN to the MN always follows the way of $CN \rightarrow MAP \rightarrow PAR \rightarrow NAR \rightarrow MN$. Under the circumstances, we obtain the following expressions:

$$t'_{New} = t_{CN-MAP} + t_{MAP-PAR} + t_{PAR-NAR} + t_{NAR-MN} = t_{New} + t_{PAR-NAR} \quad (4.23)$$

$$t_{SBU} = t_{MN-PAR} \quad (4.24)$$

$$t_{forwarding} = t_L + t_I + t_{SBU} + t'_{New} \quad (4.25)$$

In SMIPv6, the bidirectional secure tunnels are established prior to the actual

handoff, and such pre-configured tunnels support packet forwarding from the PAR to the NAR and are used to avoid packet losses. Hence, using the pre-configured tunnels, the PAR starts forwarding packets to the NAR after receiving an SBU message from the MN just before the MN disconnects from its old link. From this point of view, SMIPv6 can completely eliminate packet drops during handoff. Hence, we obtain the following equations:

$$t_{loss} = 0 \tag{4.26}$$

$$C^{p}_{SMIPv6} = \delta \times \lambda \times (t_L + t_I + t_{SBU} + t'_{New}) \tag{4.27}$$

4.5.3 Numerical results

In this section, we present some numerical results obtained from the aforementioned cost functions.

signalling cost versus L2 trigger time

To analyze the influence of L2 trigger time on the signalling cost, we define $d_{PAR-NAR} =$ 2 hops, $d_{AR-MAP} = 2$ hops, $d_{HA-MAP} = 6$ hops, $d_{CN-MAP} = 4$ hops, $d_{CN-HA} = 6$ hops. And the processing cost at each node is set as: $P_{AR} = 5$, $P_{CN} = 5$, $P_{MAP} = 10$, $P_{HA} = 20$. $\alpha = 0.2$, $\beta = 0.8$, $\kappa = 10$ and $\tau = 1$. Moreover, we define the success probability as follows [167]:

$$P_s = \frac{1}{e^{\epsilon \times t}} \tag{4.28}$$

where ϵ is a decreasing factor and t is the time taken from the occurrence of the L2 trigger to the start of the real L2 switching process.

Figure 4.14 shows the relationship between signalling costs and L2 trigger time. From this figure, we observe that MIPv6 with route optimization (RO) mode requires the most signalling cost amongst all schemes while the proposed SMIPv6 needs the least signalling costs during handoff. The average cost for SMIPv6 is 11.08; 35.68 for FMIPv6; 48.42 for F-HMIPv6; 93.00 for HMIPv6 and 252.00 for MIPv6 with RO mode. In addition, we observe that the change of the L2 trigger time has no impact on the performance of MIPv6 and HMIPv6. Yet, it results in the augmentation of signalling costs for FMIPv6, F-HMIPv6 and SMIPv6.

Figure 4.15 shows the relationship between the signalling costs and L2 trigger time when we change the value of the decreasing factor. From this figure, we observe

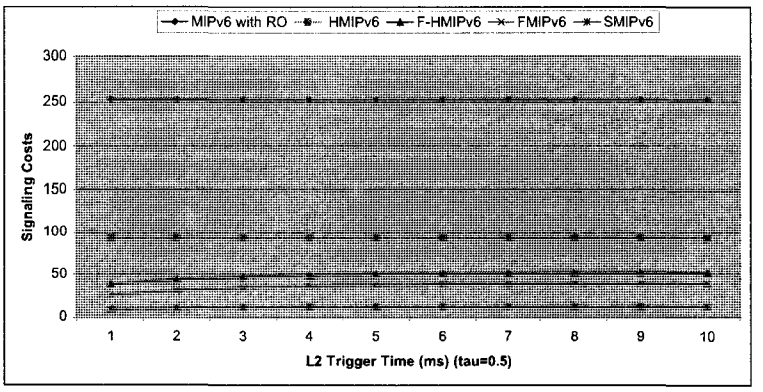

Figure 4.14 signalling Cost vs. L2 Trigger Time ($\tau = 0.5$)

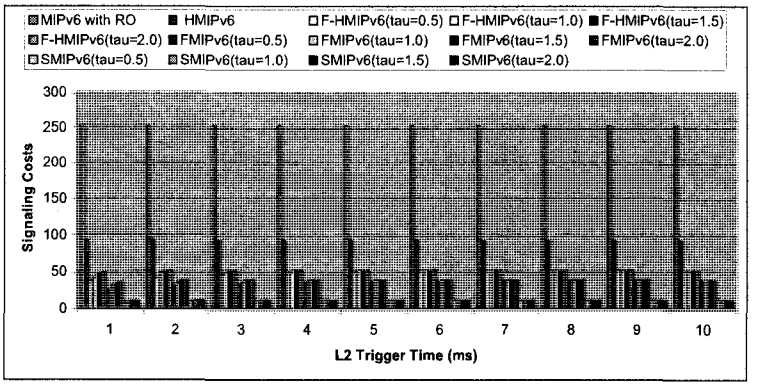

Figure 4.15 signalling Cost vs. L2 Trigger Time

that MIPv6 with route optimization (RO) mode requires the most signalling cost amongst all schemes, while the proposed SMIPv6 delivers better performance than other solutions. In addition, we observe that the change of the L2 trigger time has no impact on the performance of MIPv6 and HMIPv6. Yet, it results in the augmentation of signalling costs for FMIPv6, F-HMIPv6 and SMIPv6. However, when the L2 trigger time increases to certain value, the signalling cost curves for FMIPv6 tend to merge into a stable value. The same case is applicable for F-HMIPv6 and SMIPv6. For example, when $t \geq 9$, F-HMIPv6 (tau=1.0) merges to 52.00; when $t \geq 6$, F-HMIPv6 (tau=1.5) merges to 52.00; when $t \geq 5$, F-HMIPv6 (tau=1.5) merges to 52.00. The stable value for FMIPv6 is 39.20; 12.00 for SMIPv6. Such values are shown in Figure 4.16.

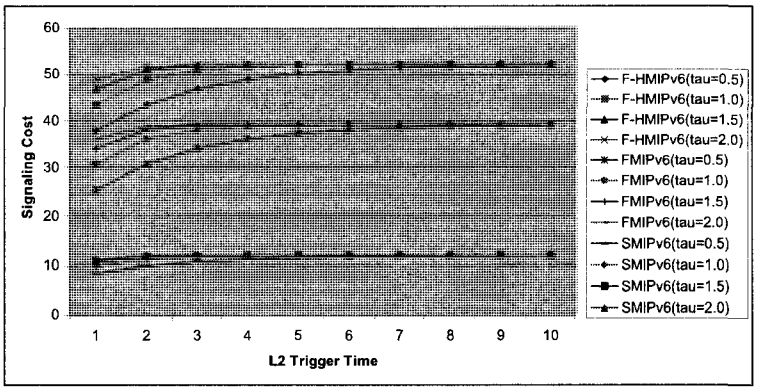

Figure 4.16 signalling Cost vs. L2 Trigger Time

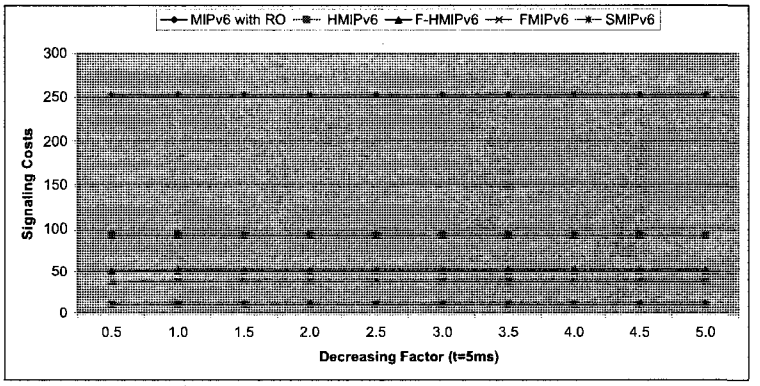

Figure 4.17 signalling Cost vs. Decreasing Factor

signalling cost versus decreasing factor

Figure 4.17 shows the relationship between the signalling costs and decreasing factor. In this analysis, the L2 trigger time is set to $5ms$. From this figure, we observe that MIPv6 with route optimization (RO) mode requires the most signalling cost amongst all schemes, while the proposed SMIPv6 needs the least signalling costs. In addition, we observe that the change of the decreasing factor has no impact on the performance of MIPv6 and HMIPv6. Yet, it results in the augmentation of signalling costs for FMIPv6, F-HMIPv6 and SMIPv6. However, when the decreasing factor increases to certain value, the signaling cost curves for FMIPv6 tend to merge into a stable value. The same case is applicable for F-HMIPv6 and SMIPv6. For example, when $\tau \geq 2.0$, F-HMIPv6 merges to 52.00; when $\tau \geq 2.0$, FMIPv6 merges to 39.20;

when $\tau \geq 1.5$, SMIPv6 merges to 12.00. We also find that the time for SMIPv6 reaches to the stable value is earlier than FMIPv6 and F-HMIPv6. In other words, the decreasing factor has less influence on SMIPv6 than on FMIPv6 and F-HMIPv6.

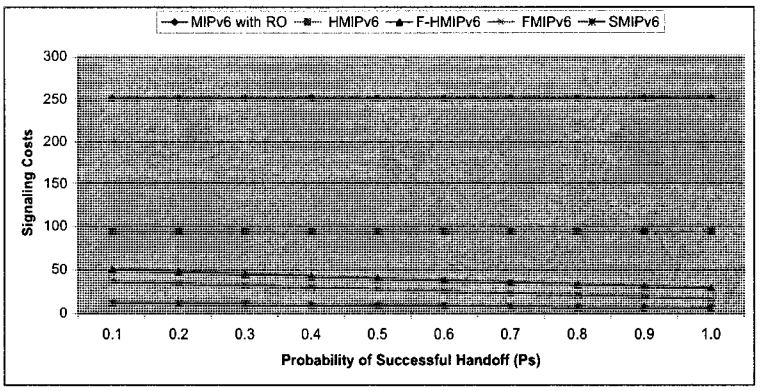

Figure 4.18 signalling Cost vs. the Probability of Successful Anticipation

signalling cost versus the probability of successful anticipation

Figure 4.18 shows the relationship between the signalling costs and probability of successful anticipation. From this figure, we observe that MIPv6 with RO mode requires the most signalling cost amongst all schemes, while the proposed SMIPv6 needs the least signalling costs. In addition, we observe that the change of the probability has no impact on the performance of MIPv6 and HMIPv6. Yet, it reduces signalling costs for FMIPv6, F-HMIPv6 and SMIPv6. The average signalling costs for SMIPv6 are 8.70; 26.55 for FMIPv6, 39.13 for F-HMIPv6; 93.00 for HMIPv6 and 252.00 for MIPv6 with RO mode.

Packet delivery cost versus L2 trigger time

To analyze the influence of t_{L2} to the packet delivery cost, we exploit the same values as in [167] with the assumption that the network topology used for analysis is symmetric. That is, $t_{CN-PAR} = t_{CN-NAR} = 150ms$, $t_{BU} = t_{New} = t_{CN-NAR} + t_{MN-NAR} = 160ms$, $t_{MN-NAR} = t_{MN-PAR} = 10ms$ and $t_{PAR-NAR} = 5ms$. And $\delta = 0.2$ and $\gamma = 0.8$; $\lambda = 1$ packets per second (pps).

Figure 4.19 shows the relationship between the packet delivery cost and L2 trigger time. From this figure, we observe that MIPv6 has the highest signaling cost while

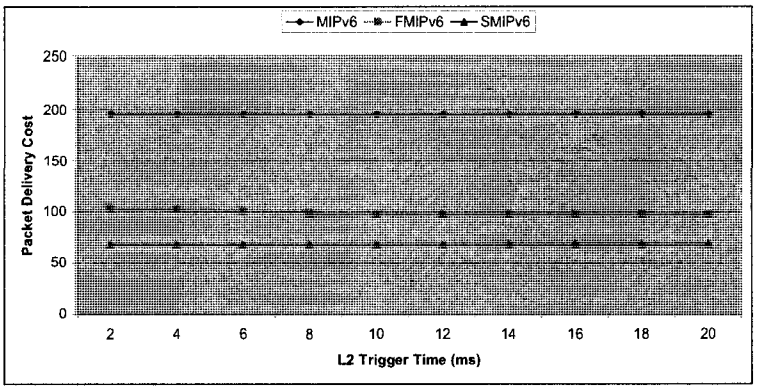

Figure 4.19 Packet Delivery Cost vs. L2 Trigger Time

SMIPv6 needs the least signalling costs. In addition, we observe that the change of L2 trigger time has no impact on the performance of MIPv6 and SMIPv6. Yet, it reduces signalling costs for FMIPv6. The average signalling costs for SMIPv6 are 68.00; 98.60 for FMIPv6 and 196.00 for MIPv6.

Packet delivery cost versus packet arrival rate

Figure 4.20 shows the relationship between the packet delivery cost and packet arrival rate. From this figure, we observe that increasing the packet arrival rate results in an augmentation of packet delivery cost. In addition, MIPv6 requires the most signalling cost while SMIPv6 deliver better performance than other approaches. The average signalling costs for SMIPv6 are 374.00; 543.10 for FMIPv6 and 1078.00 for MIPv6.

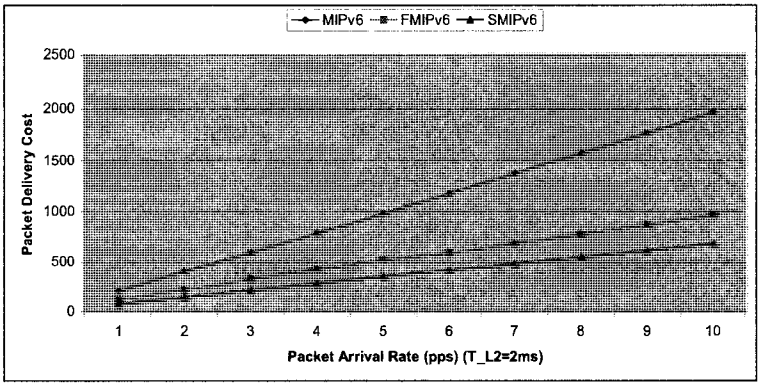

Figure 4.20 Packet Delivery Cost vs. Packet Arrival Rate

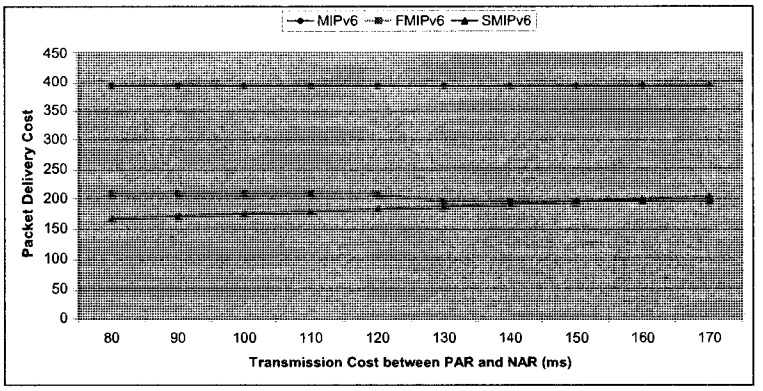

Figure 4.21 Packet Delivery Cost vs. Transmission Cost between PAR and NAR

Packet delivery cost versus transmission cost between PAR and NAR

Figure 4.21 shows the relationship between the packet delivery cost and transmission cost between the PAR and NAR. From this figure, we observe that increasing the transmission costs results in an augmentation of packet delivery cost for SMIPv6. when $t_{PAR-NAR} \geq 160ms$, SMIPv6 requires more costs than FMIPv6. In other words, SMIPv6 outperforms FMIPv6 on the condition that the transmission cost between the PAR and NAR is subjective to certain threshold. From the figure, we also find that MIPv6 requires the most costs. The average signalling costs for SMIPv6 are 184.00, 200.10 for FMIPv6, and 392.00 for MIPv6.

4.6 Analytical modeling 2

Generally, performance evaluation of mobility management schemes is based on simulation and testbed approaches [94]-[96]. Nevertheless, network scenarios for simulations vary greatly, the handoff performance comparison of the aforementioned mobility management protocols is rarely viable. Hence, in this section, we propose analytical models to analyze the performance of roaming users in IPv6-based wireless cellular networks.

4.6.1 IPv6-based cellular network architecture

We adopt the IPv6-based wireless cellular networks to evaluate the handoff performance of the above-mentioned protocols. We assume that mobile service areas are partitioned into cells of equal size. Each cell is surrounded by rings of cells [97]. And each domain is composed of n rings of the same size. We name the inmost cell "0" central cell. Cells labeled "1" constitute the first ring around cell "0", and so on. Each ring is labeled in accordance with the distance to the cell "0". To simplify the analysis, we also assume that each cell is managed by one AR. Figure 4.22 illustrates an example of a MAP domain with 3 rings.

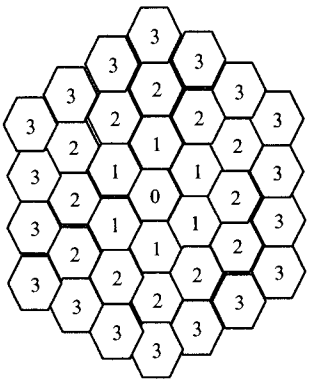

Figure 4.22 Network Topology for a MAP Domain with 3 Rings

4.6.2 Mobility models

The literature documents two common mobility models: the fluid-flow model [99]-[102] and random-walk model [97], [103]-[105]. The fluid-flow model is most appropriate for users with high mobility, few speed and direction changes while the random-walk model is best suited for pedestrian movements where mobility is confined to a limited geographical area, such as a residential neighborhood or a commercial building. Our investigation considers both models.

The random-walk model

In the random-walk mobility model, the MN's subsequent position is determined by adding a random variable with an arbitrary distribution to its previous position [97],

[103]-[105]. For an MN located in a cell of ring r, the probability that the MN moves forward to a cell of ring $(r+1)$ or backward to a cell of ring $(r-1)$ can be expressed as follows [97]:

$$p^{+}(r) = \frac{1}{3} + \frac{1}{6r} \tag{4.29}$$

$$p^{-}(r) = \frac{1}{3} - \frac{1}{6r} \tag{4.30}$$

The random-walk model can be presented by a one-dimensional Markov chain model. And the Markov chain state diagram is shown in Figure 4.23.

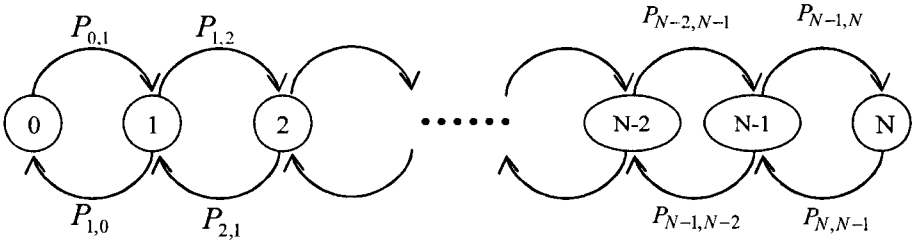

Figure 4.23 Markov Chain State Diagram for the Random-walk Model

The state r of this Markov chain is defined as the distance from the current cell where it locates the MN relative to the center cell. Furthermore, the MN is in state r if and only if it is now residing in one cell of ring r.

Assuming that each MAP controls a domain of N rings, the probability for an MN to remain in the current cell is denoted by q, the probability that the MN moves to another cell, equals $1-q$. The transition probability $P_{r,r+1}$ and $P_{r,r-1}$ represent the probability that an MN moves from its current state r to the state $(r+1)$ and $(r-1)$, respectively. They can be expressed as follows:

$$P_{r,r+1} = \begin{cases} 1-q & if r = 0 \\ (1-q) * (\frac{1}{3} + \frac{1}{6r}) & if 1 \leq r \leq N \end{cases} \tag{4.31}$$

$$P_{r,r-1} = (1-q) * (\frac{1}{3} - \frac{1}{6r}) \tag{4.32}$$

Let $\Phi_{r,N}$ be the steady-state probability of state r within a MAP domain of N rings. Using the transition probabilities in Equations 4.31 and 4.32, $\Phi_{r,R}$ is expressed

as:

$$\Phi_{r,N} = \Phi_{0,N} \prod_{i=0}^{r-1} \frac{P_{i,i+1}}{P_{i+1,i}} \tag{4.33}$$

As $\sum_{r=0}^{N} \Phi_{r,N} = 1$, $\Phi_{0,N}$ can be expressed as:

$$\Phi_{0,N} = \frac{1}{\sum_{r=1}^{N} \prod_{i=0}^{r-1} \frac{P_{i,i+1}}{P_{i+1,i}}} \tag{4.34}$$

We assume that each cell is controlled by an AP which integrates AR functionalities, and each MAP controls a domain of N rings, the probability that an MN performs an inter-domain location update is expressed as:

$$P = \Phi_{N,N} \times P_{N,N+1} \tag{4.35}$$

Where $\Phi_{N,N}$ denotes the steady-state probability of state N, $P_{N,N+1}$ denotes the probability that the MN moves from a cell in ring N to a cell in ring $(N+1)$.

The fluid-flow model

Using the fluid-flow model, the MN's directional movement is distributed uniformly in the range of $(0, 2\pi)$. Let v denote the MN's average speed (m/s), L_c, L_d the cell and domain perimeters (m), S_c and S_d represent the area of a cell and a domain (m^2), and R the cell radius (m). Within a MAP domain of N rings, the cell crossing rate R_c and domain crossing rate R_d are given as follows:

$$R_c = \frac{v \times L_c}{\pi \times S_c} = \frac{v \times 6R}{\pi \times 2.6R^2} = \frac{6v}{\pi \times 2.6R} \tag{4.36}$$

$$R_d = \frac{v \times L_d}{\pi \times S_d} = \frac{v \times (12N+6)}{\pi \times [3N \times (N+1)+1] \times 2.6R} \tag{4.37}$$

4.6.3 Handoff related signalling overhead function

We assume that HMIPv6, F-HMIPv6, FMIPv6 and SMIPv6 support route optimization (RO) and only a pair of messages (neighbor solicitation and neighbor advertisement) exchanged for the DAD process. Both MN and CN processing costs are ignored

during analysis. The signalling overhead functions for MIPv6 with tunnel mode, MIPv6 with RO mode, intra-domain HMIPv6 [8], inter-domain HMIPv6, predictive FMIPv6 [10], reactive FMIPv6 [10], intra-domain F-HMIPv6 [17] [18], inter-domain F-HMIPv6, predictive SMIPv6 and reactive SMIPv6 are listed as follows:

$$S_{MIPv6-tunnel} = 6\kappa + 2\tau \times d_{AR-HA} \tag{4.38}$$

$$S_{MIPv6-RO} = S_{MIPv6-tunnel} + N_{CN} \times [5\kappa + 2\tau \times (d_{AR-HA} + d_{HA-CN}) + 3\tau \times d_{AR-CN}] \tag{4.39}$$

$$S_{intra-HMIPv6} = 6\kappa + 4\tau \times d_{AR-MAP} \tag{4.40}$$

$$S_{inter-HMIPv6} = S_{intra-HMIPv6} + S_{MIPv6-RO} \tag{4.41}$$

$$S_{P-FMIPv6} = 5\kappa + 5\tau \times d_{PAR-NAR} \tag{4.42}$$

$$S_{R-FMIPv6} = 3\kappa + 4\tau \times d_{PAR-NAR} \tag{4.43}$$

$$S_{intra-FHMIPv6} = 7\kappa + 13\tau \times d_{AR-MAP} \tag{4.44}$$

$$S_{inter-FHMIPv6} = S_{intra-FHMIPv6} + S_{MIPv6-RO} \tag{4.45}$$

$$S_{P-SMIPv6} = 2\kappa \tag{4.46}$$

$$S_{R-SMIPv6} = \kappa \tag{4.47}$$

Where κ and τ represent the unit transmission cost in a wireless and wired link, respectively; d_{x-y} the hop distance between network entities x and y, and N_{CN} indicates the number of CNs.

Equation 4.38 implies that 6 messages (RS/RA, NS/NA, BU/BA) are exchanged between the MN and AR via radio link during handover, and the signalling cost for each message is represented by κ. In addition, 2 messages (BU/BA) are exchanged between the AR and HA via wired link, the signalling cost for each message is presented by $\tau \times d_{AR-HA}$. The same principles apply to other equations.

4.6.4 Handoff signalling costs using the random-walk model

The handoff related signalling cost function for MIPv6 with tunnel mode, MIPv6 with route optimization (RO) mode, predictive FMIPv6 (P-FMIPv6), reactive FMIPv6 (R-FMIPv6), F-HMIPv6, predictive SMIPv6 (P-SMIPv6) and reactive SMIPv6 (R-

SMIPv6) is expressed as follows:

$$C^l_{MIPv6-tunnel} = \frac{S_{MIPv6-tunnel} \times (1-q)}{E(T)} \tag{4.48}$$

$$C^l_{MIPv6-RO} = \frac{S_{MIPv6-RO} \times (1-q)}{E(T)} \tag{4.49}$$

$$C^l_{HMIPv6} = \frac{P \times S_{inter-HMIPv6} + (1-P) \times S_{intra-HMIPv6}}{E(T)} \tag{4.50}$$

$$C^l_{P-FMIPv6} = \frac{S_{P-FMIPv6} \times (1-q)}{E(T)} \tag{4.51}$$

$$C^l_{R-FMIPv6} = \frac{S_{R-FMIPv6} \times (1-q)}{E(T)} \tag{4.52}$$

$$C^l_{FHMIPv6} = \frac{P \times S_{inter-FHMIPv6} + (1-P) \times S_{intra-FHMIPv6}}{E(T)} \tag{4.53}$$

$$C^l_{P-SMIPv6} = \frac{S_{P-SMIPv6} \times (1-q)}{E(T)} \tag{4.54}$$

$$C^l_{R-SMIPv6} = \frac{S_{R-SMIPv6} \times (1-q)}{E(T)} \tag{4.55}$$

Where q denotes the probability that an MN remains in its current cell, $E(T)$ the average cell residence time and P the probability that an MN performs an inter-domain handoff.

4.6.5 Handoff signalling costs using the fluid-flow model

The handoff related signalling cost function for MIPv6 with tunnel mode, MIPv6 with RO mode, P-FMIPv6, R-FMIPv6, P-SMIPv6 and R-SMIPv6 is expressed as follows:

$$C^l_{MIPv6-tunnel} = R_c \times S_{MIPv6-tunnel} \times (1-q) \tag{4.56}$$

$$C^l_{MIPv6-RO} = R_c \times S_{MIPv6-RO} \times (1-q) \tag{4.57}$$

$$C^l_{HMIPv6} = \frac{R_d \times P \times S_{inter-HMIPv6} + (N_{AR}R_c - R_d) \times S_{intra-HMIPv6} \times (1-P)}{N_{AR}} \tag{4.58}$$

$$C^l_{P-FMIPv6} = R_c \times S_{P-FMIPv6} \times (1-q) \tag{4.59}$$

$$C^{l}_{R-FMIPv6} = R_c \times S_{R-FMIPv6} \times (1-q) \tag{4.60}$$

$$C^{l}_{FHMIPv6} = \frac{R_d \times P \times S_{inter-FHMIPv6} + (N_{AR}R_c - R_d) \times S_{intra-FHMIPv6} \times (1-P)}{N_{AR}} \tag{4.61}$$

$$C^{l}_{P-SMIPv6} = R_c \times S_{P-SMIPv6} \times (1-q) \tag{4.62}$$

$$C^{l}_{R-SMIPv6} = R_c \times S_{R-SMIPv6} \times (1-q) \tag{4.63}$$

Where R_c and R_d denote cell and domain crossing rates, q denotes the probability that an MN remains in its current cell, N_{AR} the number of ARs in a domain and P the probability that an MN performs an inter-domain handoff.

4.6.6 Packet delivery costs

Packet delivery costs per session are defined as the costs of delivering a session from a CN to an MN, including all nodes' processing costs and link transmission costs from the CN to MN. Assuming that HMIPv6, FMIPv6, F-HMIPv6 and SMIPv6 support route optimization (RO), the packet delivery cost function for each handoff management scheme is given as follows:

$$C^{p}_{MIPv6-tunnel} = P^{'}_{HA} + \kappa \times \lambda_s + \tau \times \lambda_s \times (d_{CN-HA} + d_{HA-AR}) \tag{4.64}$$

$$C^{p}_{MIPv6-RO} = P_{HA} + P_{CN} + \kappa \times \lambda_s + \tau \times [\lambda_p \times (d_{CN-HA} + d_{HA-AR}) + (\lambda_s - \lambda_p) \times d_{CN-AR}] \tag{4.65}$$

$$C^{p}_{HMIPv6} = P_{MAP} + C^{p}_{MIPv6-RO} \tag{4.66}$$

$$C^{p}_{FMIPv6} = P_{AR} + C^{p}_{MIPv6-RO} + \tau \times \lambda_s \times d_{PAR-NAR} \tag{4.67}$$

$$C^{p}_{F-HMIPv6} = C^{p}_{HMIPv6} \tag{4.68}$$

$$C^{r}_{SMIPv6} = C^{p}_{FMIPv6} \tag{4.69}$$

Where λ_s denotes the session arrival rate (packets per second), λ_p represents the average packet arrival rate (packets per second), P_z the processing cost at network entity z, d_{x-y} the hop distance between network entities x and y , κ and τ denote the unit transmission cost in a wireless and wired links, respectively.

In HMIPv6 and F-HMIPv6, each MAP maintains a binding cache table for translation between MNs' regional CoAs and their on-link local CoAs, just as the HA binds

MNs' HoAs to their CoAs. All packets addressed to an MN are intercepted by the MAP and tunnelled to the MN's new on-link local CoA. Hence, MAP processing cost includes lookup and routing costs. Lookup cost is proportional to the size of binding cache table, thus proportional to the number of MNs in a MAP domain. In addition, routing cost is proportional to the logarithm of the number of ARs in a MAP domain [104]. Therefore, the MAP processing cost can be further expressed as:

$$P_{MAP} = \lambda_s \times (\alpha \times N_{AR} \times E_{MN} + \beta \times \log_2 N_{AR}) \tag{4.70}$$

Where λ_s denotes the session arrival rate (packets per second), α is a proportionality factor that shows the relationship between the MAP's lookup cost and size of the binding cache table, β is a weighting factor that indicates the relationship between the MAP routing cost and number of ARs within a MAP domain, E_{MN} the average quantity of MNs in a cell and N_{AR} the number of ARs in a MAP domain.

Under the RO mode, only the first packet of a session is transmitted to the HA to detect whether the MN is away from its home network. All successive packets of the session are routed directly to the MN's new location. Hence, HA processing cost for RO mode is expressed as:

$$P_{HA} = \lambda_p \times \theta_{HA} \tag{4.71}$$

Where λ_p denotes arrival rate of the first packet of a session, which is assumed to be the average packet arrival rate (packets per second). θ_{HA} indicates the unit packet processing cost at an HA.

Under the tunnel mode, all packets in a session is handled by the HA. Hence, the HA processing cost for tunnel mode is given by:

$$P'_{HA} = \lambda_s \times \theta_{HA} \tag{4.72}$$

In FMIPv6, the PAR binds MNs' PCoAs to their NCoAs, so the PAR processing cost mainly comprises the lookup costs for searching its binding cache table, which is proportional to the number of MNs served by the PAR. Assuming that the processing costs for each AR is identical, the PAR processing costs can be expressed as:

$$P_{PAR} = P_{AR} = \lambda_s \times (\epsilon \times E_{MN}) \tag{4.73}$$

Where λ_s denotes the session arrival rate (packets per second), ϵ is a weighting factor

showing the relationship between the PAR's lookup cost and size of its binding cache table and E_{MN} the average number of MNs in a cell.

The random-walk model

The number of MNs in a MAP domain of N rings can be expressed as:

$$\Gamma = N_{AR} \times E(MN) = [3N \times (N+1) + 1] \times E(MN) \tag{4.74}$$

Where N_{AR} denotes the number of ARs in a domain, $E(MN)$ the average number of MNs served in a cell.

The fluid-flow model

The number of MNs in a MAP domain of N rings can be expressed as:

$$\Gamma = \rho \times A_d = \rho \times N_{AR} \times A_c = \rho \times [3N \times (N+1) + 1] \times 2.6R^2 \tag{4.75}$$

Where ρ denotes user density in a cell ($/m^2$), in terms of the number of MNs per square meter, A_c and A_d represent the area of a cell and a domain, respectively, N_{AR} the quantity of ARs in a domain of N rings and R the cell radius (m).

4.6.7 Total cost

The total cost is defined as the sum of the handoff related signalling costs and the packet delivery costs.

Total cost using the random-walk model

The total cost function for MIPv6 with tunnel mode, MIPv6 with route optimization (RO) mode, predictive FMIPv6 (P-FMIPv6), reactive FMIPv6 (R-FMIPv6), F-HMIPv6, predictive SMIPv6 (P-SMIPv6) and reactive SMIPv6 (R-SMIPv6) is expressed as follows:

$$C^t_{MIPv6-tunnel} = \frac{S_{MIPv6-tunnel} \times (1-q)}{E(T)} + C^p_{MIPv6-tunnel} \tag{4.76}$$

$$C^t_{MIPv6-RO} = \frac{S_{MIPv6-RO} \times (1-q)}{E(T)} + C^p_{MIPv6-RO} \tag{4.77}$$

$$C^t_{HMIPv6} = \frac{P \times S_{inter-HMIPv6} + (1-P) \times S_{intra-HMIPv6}}{E(T)} + C^p_{HMIPv6} \tag{4.78}$$

$$C^t_{P-FMIPv6} = \frac{S_{P-FMIPv6} \times (1-q)}{E(T)} + C^p_{FMIPv6} \tag{4.79}$$

$$C^t_{R-FMIPv6} = \frac{S_{R-FMIPv6} \times (1-q)}{E(T)} + C^p_{FMIPv6} \tag{4.80}$$

$$C^t_{FHMIPv6} = \frac{P \times S_{inter-FHMIPv6} + (1-P) \times S_{intra-FHMIPv6}}{E(T)} + C^p_{FHMIPv6} \tag{4.81}$$

$$C^t_{P-SMIPv6} = \frac{S_{P-SMIPv6} \times (1-q)}{E(T)} + C^p_{SMIPv6} \tag{4.82}$$

$$C^t_{R-SMIPv6} = \frac{S_{R-SMIPv6} \times (1-q)}{E(T)} + C^p_{SMIPv6} \tag{4.83}$$

Where q denotes the probability that an MN remains in its current cell, $E(T)$ the average cell residence time, P the probability that an MN performs an inter-domain handoff, and $C^p_{MIPv6-tunnel}$ is given by Equation (4.64); $C^p_{MIPv6-RO}$ is expressed by Equation (4.65); C^p_{HMIPv6} is expressed by Equation (4.66); C^p_{FMIPv6} is expressed by Equation (4.67); $C^p_{FHMIPv6}$ is given by Equation (4.68) and C^p_{SMIPv6} is expressed by Equation (4.69).

Total cost using the fluid-flow model

The total cost function for MIPv6 with tunnel mode, MIPv6 with route optimization (RO) mode, predictive FMIPv6 (P-FMIPv6), reactive FMIPv6 (R-FMIPv6), F-HMIPv6, predictive SMIPv6 (P-SMIPv6) and reactive SMIPv6 (R-SMIPv6) is expressed as follows:

$$C^t_{MIPv6-tunnel} = R_c \times S_{MIPv6-tunnel} \times (1-q) + C^p_{MIPv6-tunnel} \tag{4.84}$$

$$C^t_{MIPv6-RO} = R_c \times S_{MIPv6-RO} \times (1-q) + C^p_{MIPv6-RO} \tag{4.85}$$

$$C^t_{HMIPv6} = C^l_{HMIPv6} + C^p_{HMIPv6} \tag{4.86}$$

$$C^t_{P-FMIPv6} = R_c \times S_{P-FMIPv6} \times (1-q) + C^p_{FMIPv6} \tag{4.87}$$

$$C^t_{R-FMIPv6} = R_c \times S_{R-FMIPv6} \times (1-q) + C^p_{FMIPv6} \tag{4.88}$$

$$C^t_{FHMIPv6} = C^l_{FHMIPv6} + C^p_{FHMIPv6} \tag{4.89}$$

$$C^t_{P-SMIPv6} = R_c \times S_{P-SMIPv6} \times (1-q) + C^p_{SMIPv6} \tag{4.90}$$

$$C^t_{R-SMIPv6} = R_c \times S_{R-SMIPv6} \times (1-q) + C^p_{SMIPv6} \tag{4.91}$$

Where R_c and R_d denote cell and domain crossing rates, q denotes the probability that an MN remains in its current cell, N_{AR} the number of ARs in a domain, P the probability that an MN performs an inter-domain handoff, C^l_{HMIPv6} is given by Equation (4.58); $C^l_{FHMIPv6}$ is expressed by Equation (4.61); $C^p_{MIPv6-tunnel}$ is given by Equation (4.64); $C^p_{MIPv6-RO}$ is expressed by Equation (4.65); C^p_{HMIPv6} is expressed by Equation (4.66); C^p_{FMIPv6} is expressed by Equation (4.67); $C^p_{FHMIPv6}$ is given by Equation (4.68) and C^p_{SMIPv6} is expressed by Equation (4.69).

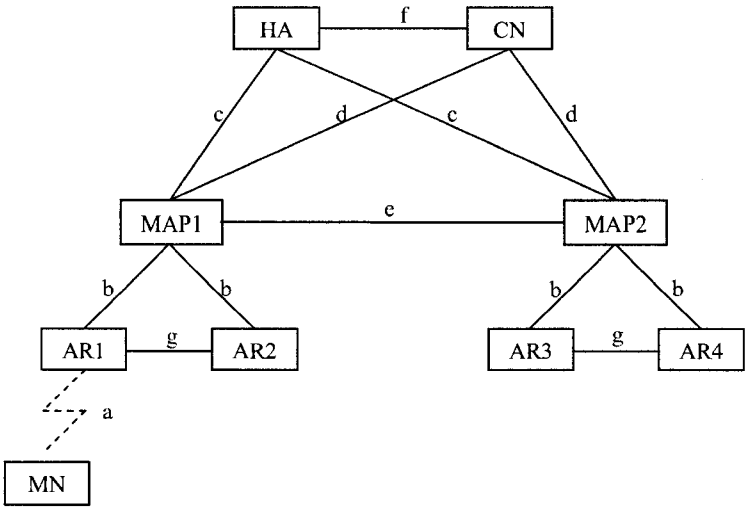

Figure 4.24 Network Topology Used for Performance Analysis

4.6.8 Numerical results

This section analyzes the impact of various wireless system parameters on the above-mentioned costs. The parameter values are taken from [99], [104], [106], [107], i.e. $\alpha = 0.1$ and $\beta = 0.2$, $\lambda_s = 1$, $\lambda_p = 0.1$, $\theta_{HA} = 20$, $\tau = 1$, $\kappa = 2$, $N_{CN} = 2$, $L_c = 120m$, the network topology is shown in Figure 4.24. Additionally, we add the

value of $\epsilon = 0.1$, $R = 20m$. The hop distance between different domains is assumed to be identical, i.e. $d_{HA-CN} = f = 6$, $d_{CN-MAP} = d = 4$, $d_{HA-MAP} = c = 6$, $d_{AR-MAP} = b = 2$, $d_{AR1-AR2} = d_{PAR-NAR} = 2$. And all links are assumed to be full-duplex in terms of capacity and delay.

Handoff related signalling costs versus cell residence time

Figure 4.25 shows the relationship between handoff related signalling costs and cell residence time for $q = 0.2$, using the random-walk model. This figure shows dynamic MNs eager to move to another cell. We observe that longer cell residence time yields lower signalling costs. This is to be expected given that fewer handoffs are required as MNs remain longer in their current cells. Additionally, our proposed SMIPv6 schemes deliver better performance than other handoff schemes while MIPv6 with RO mode requires the most handoff signalling cost. The mean costs are 32.80 for MIPv6 with RO, 25.77 for F-HMIPv6, 19.92 for HMIPv6, 6.56 for MIPv6 with tunnel mode, 4.69 for predictive FMIPv6 (P-FMIPv6) and 3.28 for reactive FMIPv6 (R-FMIPv6), 0.94 for predictive SMIPv6 (P-SMIPv6) and 0.47 for reactive SMIPv6 (R-FMIPv6).

Figure 4.26 shows the relationship between handoff related signalling costs and cell residence time for $q = 0.8$, using the random-walk model. This figure shows static MNs, highly likely to remain in their current cells. From the figure, we find that longer cell residence time yields lower signaling costs. This is to be expected given that fewer handoffs are required as MNs remain longer in their current cells. Additionally, results indicate that dynamic MNs require more handoff signalling costs than static MNs. Moreover, our proposed SMIPv6 schemes provide better performance than other handoff schemes while F-HMIPv6 requires the most handoff signalling cost. The mean costs are 15.23 for F-HMIPv6, 9.37 for HMIPv6, 8.20 for MIPv6 with RO, 1.64 for MIPv6 with tunnel mode, 1.17 for predictive FMIPv6 (P-FMIPv6) and 0.82 for reactive FMIPv6 (R-FMIPv6), 0.23 for predictive SMIPv6 (P-SMIPv6) and 0.12 for reactive SMIPv6 (R-SMIPv6).

Handoff related signalling costs versus user velocity

Figure 4.27 shows the relationship between handoff related signalling costs and user velocity for a MAP domain with one ring, using the fluid-flow model. Here the probability that the MN remains at its current cell is set to 0.2. Handoff related

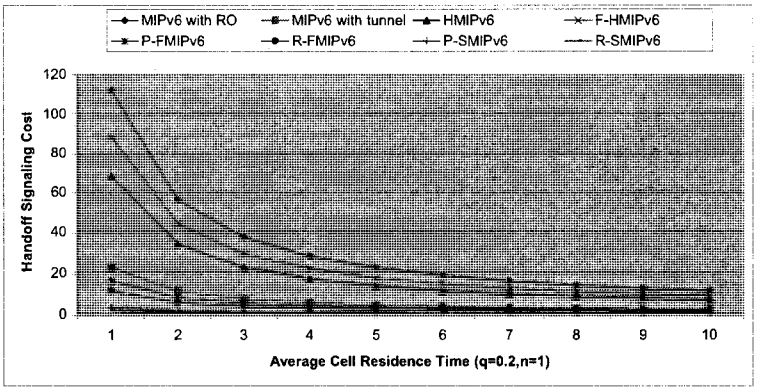

Figure 4.25 Handoff signalling Costs vs. Cell Residence Time ($q = 0.2$)

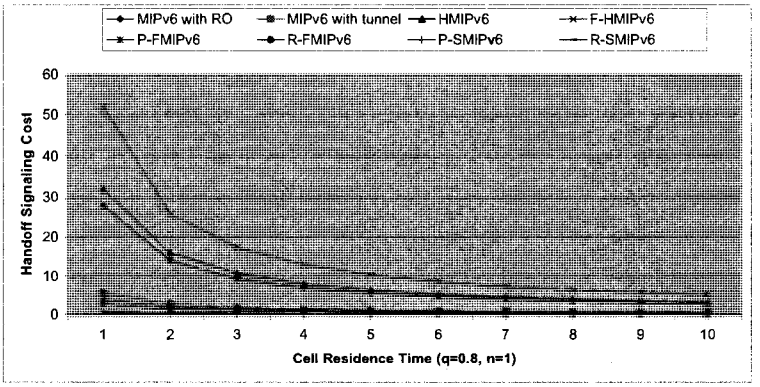

Figure 4.26 Handoff signalling Costs vs. Cell Residence Time ($q = 0.8$)

signalling costs increase linearly as MNs' average velocity augments. Since MNs with a higher mean velocity are more likely to cross a cell and a domain, resulting in higher signalling costs. This figure shows that MIPv6 with RO engenders the most exorbitant cost, its signalling costs rising to 113.12, on average. In comparison, F-HMIPv6 climbs to 28.74, 22.62 for MIPv6 with tunnel mode, and 16.16 for predictive FMIPv6 (P-FMIPv6), 15.85 for HMIPv6, 11.31 for reactive FMIPv6 (R-FMIPv6), and 3.23 for predictive SMIPv6 (P-SMIPv6) and 1.62 for reactive SMIPv6 (R-SMIPv6).

Figure 4.28 show the relationship between location update costs and user velocity for a MAP domain with four rings, using the fluid-flow model. The probability that the MN remains at its current cell is set to 0.2. Handoff related signalling costs

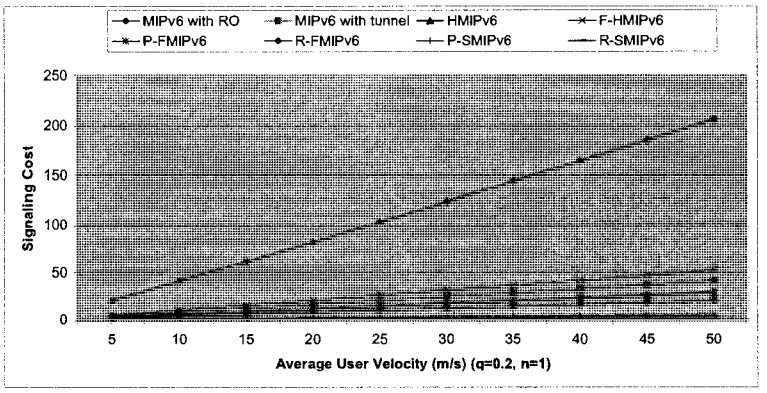

Figure 4.27 Handoff signalling Costs vs. User Velocity ($N = 1$)

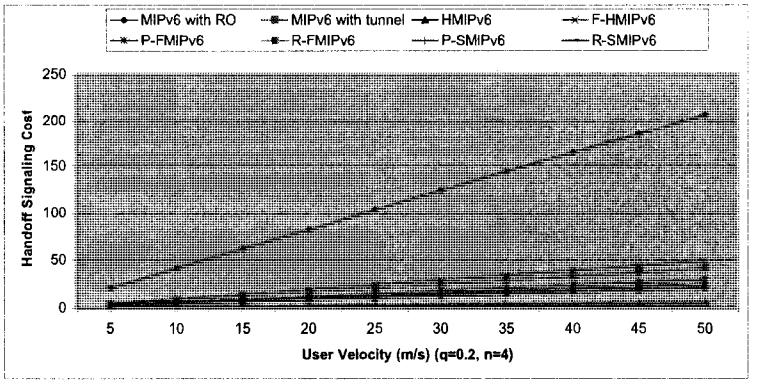

Figure 4.28 Handoff signalling Costs vs. User Velocity ($N = 4$)

increase linearly as MNs' average velocity augments. In addition, the domain size increase has no impact on the performance of MIPv6, FMIPv6 and SMIPv6, but leads to the reduced signalling costs for HMIPv6 and F-HMIPv6. This can be explained by the fact that MNs that roam in larger domains are less likely to perform handoffs. In this case, F-HMIPv6 descends to 26.64 while HMIPv6 descends to 13.38, on average.

signalling costs versus cell residence time

Figure 4.29 shows the relationship between handoff related signalling costs and cell residence time for $N = 1$ and $q = 0.4$, using the random-walk model. In this analysis, we focus on the effect of changing domain size on the signalling costs. We observe that

longer cell residence time yields lower handoff signalling costs. This is to be expected given that fewer handoffs are required as MNs remain longer in their current cells. Additionally, our proposed SMIPv6 schemes deliver better performance than other handoff schemes while MIPv6 with RO mode requires the most handoff signalling cost. The mean costs are 24.60 for MIPv6 with RO, 22.26 for F-HMIPv6, 16.40 for HMIPv6, 4.92 for MIPv6 with tunnel mode, 3.51 for predictive FMIPv6 (P-FMIPv6) and 2.46 for reactive FMIPv6 (R-FMIPv6), 0.70 for predictive SMIPv6 (P-SMIPv6) and 0.35 for reactive SMIPv6 (R-SMIPv6).

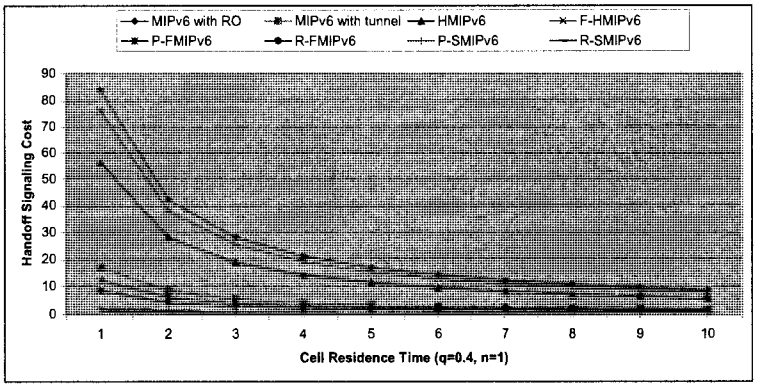

Figure 4.29 Handoff signalling Costs vs. Cell Residence Time ($N = 1$)

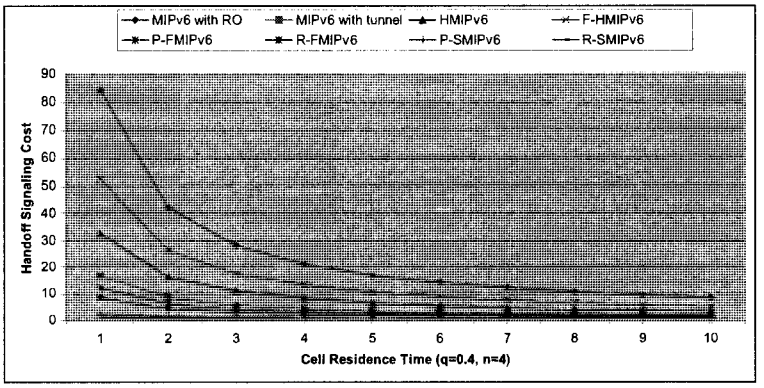

Figure 4.30 Handoff signalling Costs vs. Cell Residence Time ($N = 4$)

Figure 4.30 shows the relationship between handoff related signalling costs and cell

residence time for $N = 4$ and $q = 0.4$, using the random-walk model. We observe that longer cell residence time yields lower location update costs. This is to be expected given that fewer location updates are required as MNs remain longer in their current cells. Moreover, the increasing of MAP domain size has no impact on the performance of MIPv6, FMIPv6 and SMIPv6. Yet, results in significant reduced signalling cost for HMIPv6 and F-HMIPv6. We explain this by the fact that an MN moving in a larger domain is less likely to perform inter-domain handoffs. Additionally, our proposed SMIPv6 schemes deliver better performance than other handoff schemes while MIPv6 with RO mode requires the most handoff signalling cost. The mean costs are 24.60 for MIPv6 with RO mode, 15.35 for F-HMIPv6, 9.49 for HMIPv6, 4.92 for MIPv6 with tunnel mode, 3.51 for predictive FMIPv6 (P-FMIPv6) and 2.46 for reactive FMIPv6 (R-FMIPv6), 0.70 for predictive SMIPv6 (P-SMIPv6) and 0.35 for reactive SMIPv6 (R-SMIPv6).

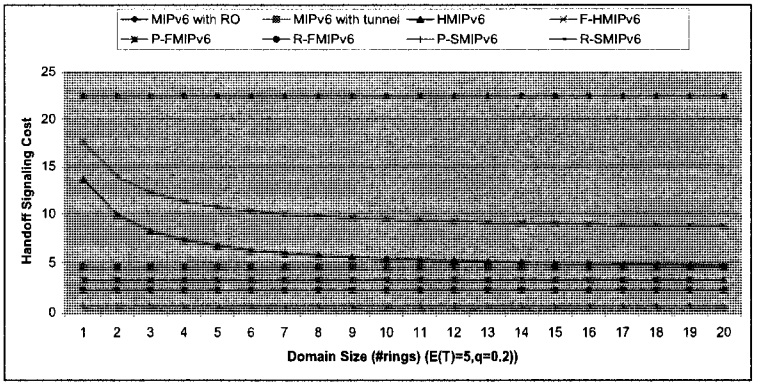

Figure 4.31 Handoff signalling Costs vs. Domain Size ($q = 0.2$)

Handoff related signalling costs versus domain size

Figure 4.31 shows handoff related signalling costs compared to domain size for $q = 0.2$ and $E(T) = 5s$ under the random-walk model. As domain size increases, handoff related signalling costs are largely reduced for both HMIPv6 and F-HMIPv6, although the increasing of domain size does not affect the performance of MIPv6, FMIPv6 and SMIPv6. The average signalling cost for MIPv6 with RO mode is 22.40; 10.22 for F-HMIPv6, 6.22 for HMIPv6, 4.48 for MIPv6 with tunnel mode, 3.20 for Predictive

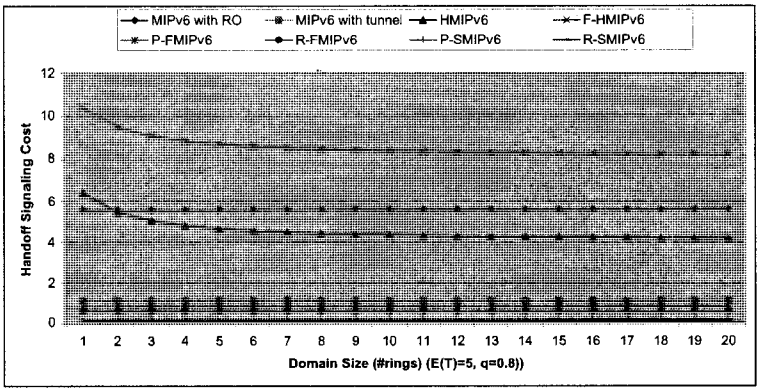

Figure 4.32 Handoff signalling Costs vs. Domain Size ($q = 0.8$)

FMIPv6 (P-FMIPv6) and 2.24 for Reactive FMIPv6 (R-FMIPv6), 0.64 for Predictive SMIPv6 (P-SMIPv6) and 0.32 for Reactive SMIPv6 (R-SMIPv6). In addition, our proposed SMIPv6 schemes yield better performance than other approaches.

Figure 4.32 also shows handoff related signalling costs compared to domain size for $q = 0.8$ and $E(T) = 5s$ under the random-walk model. The increasing of domain size results in drastically reduced handoff related signalling costs for both HMIPv6 and F-HMIPv6, but has no impact on the performance of MIPv6, FMIPv6 and SMIPv6. Figure 4.32 shows that when the domain size counts more than 1 ring, HMIPv6 requires less signalling costs than MIPv6 with RO mode. Figure 4.31 presents a scenario where dynamic MNs are eager to perform handoffs while Figure 4.32 shows static MNs, which are highly likely to remain in their current cells. With static MNs, numerical results show that F-HMIPv6 requires more handoff signalling costs than other solutions. The average signaling cost for F-HMIPv6 is 8.56; 5.60 for MIPv6 with RO mode, 4.56 for HMIPv6, 1.12 for MIPv6 with tunnel mode, 0.80 for predictive FMIPv6 (P-FMIPv6) and 0.56 for reactive FMIPv6 (R-FMIPv6), 0.16 for predictive SMIPv6 (P-SMIPv6) and 0.08 for reactive SMIPv6 (R-SMIPv6). In addition, this figure also shows that our proposed SMIPv6 schemes yield better performance than other approaches.

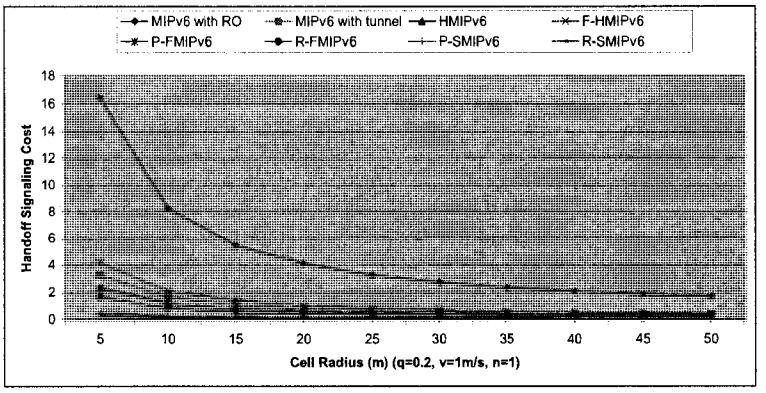

Figure 4.33 Handoff Signalling Costs vs. Cell Radius ($N = 1$)

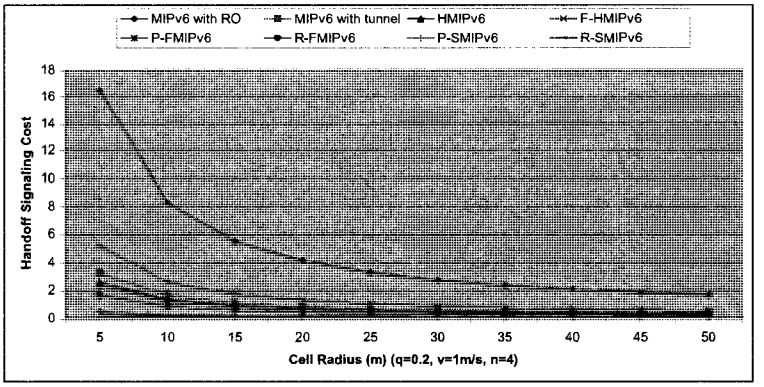

Figure 4.34 Handoff Signalling Costs vs. Cell Radius ($N = 4$)

Handoff signalling costs versus cell sizes

Figure 4.33 shows the relationship between handoff related signalling costs and cell sizes for a MAP domain with one ring and $q = 0.2$ and $v = 1m/s$ under the fluid-flow model. As cell radius increases, handoff related signalling costs reduces. We explain this by the fact that MNs roaming in a larger cell are less likely to perform handoffs. The average signalling cost for MIPv6 with RO mode is 4.82; 1.22 for F-HMIPv6, 0.96 for MIPv6 with tunnel mode, 0.69 for predictive FMIPv6 (P-FMIPv6), 0.68 for HMIPv6, and 0.48 for reactive FMIPv6 (R-FMIPv6), 0.14 for predictive SMIPv6 (P-SMIPv6) and 0.07 for reactive SMIPv6 (R-SMIPv6). In addition, our proposed SMIPv6 schemes yield better performance than other approaches.

Figure 4.34 shows the relationship between handoff related signalling costs and cell radius for a MAP domain with 4 rings and $q = 0.2$ and $v = 1m/s$ under the fluid-flow model. As cell radius increases, handoff related signalling costs reduces. We explain this by the fact that MNs roaming in a small cell are more likely to perform handoffs. The average signalling cost for MIPv6 with RO mode is 4.82; 1.52 for F-HMIPv6, 0.96 for MIPv6 with tunnel mode, 0.76 for HMIPv6, 0.69 for Predictive FMIPv6 (P-FMIPv6), and 0.48 for Reactive FMIPv6 (R-FMIPv6), 0.14 for Predictive SMIPv6 (P-SMIPv6) and 0.07 for Reactive SMIPv6 (R-SMIPv6). In addition, our proposed SMIPv6 schemes provide better performance than other approaches.

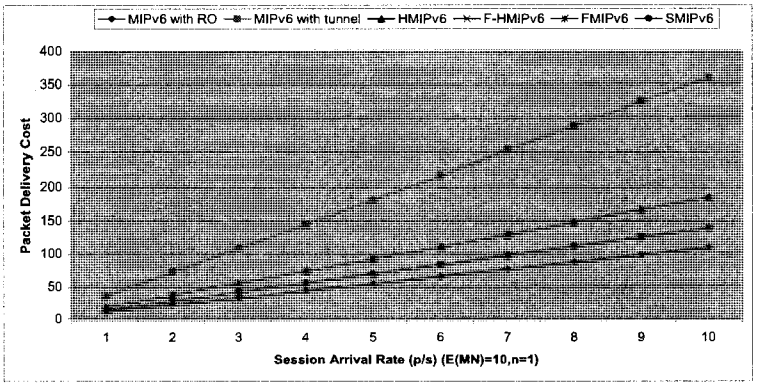

Figure 4.35 Packet Delivery Costs vs. Session Arrival Rate ($N = 1$)

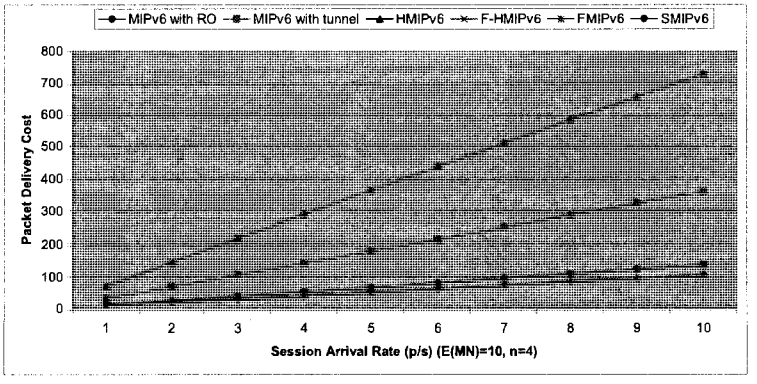

Figure 4.36 Packet Delivery Costs vs. Session Arrival Rate ($N = 4$)

Packet delivery costs versus session arrival rates

Figure 4.35 illustrates the relationship between packet delivery costs and session arrival rates for a MAP domain with 1 ring. Generally, the higher the session arrival rate, the higher the packet delivery costs. From the figure, we observe that MIPv6 with tunnel mode requires the highest costs amongst all schemes, since all of the session packets must cross a triangular path via the HA, whose steep processing costs are detrimental. Furthermore, MIPv6 with RO mode provides better performance than other approaches, since all the packets (except the first one) in a session are delivered to the MN through a direct path, and there is no additional processing cost at the MAP neither at the AR. HMIPv6 and F-HMIPv6 deliver identical performance, as do FMIPv6 and SMIPv6. The mean costs are 198.00 for MIPv6 with tunnel mode, 100.99 for F-HMIPv6 and HMIPv6, and 75.90 for FMIPv6 and SMIPv6, 59.40 for MIPv6 with RO mode.

Figure 4.36 shows the relationship between packet delivery costs and session arrival rates for a MAP domain with 4 rings. Generally, the higher the session arrival rate, the higher the packet delivery costs. MIPv6 with tunnel mode requires the highest costs, since all of the session packets must cross a triangular path via the HA, whose steep processing costs are detrimental. In addition, increasing the domain size results in higher packet delivery costs for both HMIPv6 and F-HMIPv6. This is because increasing the number of ARs in a domain causes higher MAP processing costs. Furthermore, MIPv6 with RO mode provides better performance than other approaches, since all the packets (except the first one) in a session are delivered to the MN through a direct path, and there is no additional processing cost at the MAP neither at the AR. HMIPv6 and F-HMIPv6 deliver identical performance, as do FMIPv6 and SMIPv6. The mean costs are 401.42 for F-HMIPv6 and HMIPv6, 198.00 for MIPv6 with tunnel mode, and 75.90 for FMIPv6 and SMIPv6, 59.40 for MIPv6 with RO mode. In addition, increasing of MAP domain size has no impact on the performance of MIPv6, FMIPv6 and SMIPv6. Yet, it results in the augmentation of packet delivery costs for HMIPv6 and F-HMIPv6. Because the lookup costs and routing costs increase with the augmentation of domain size.

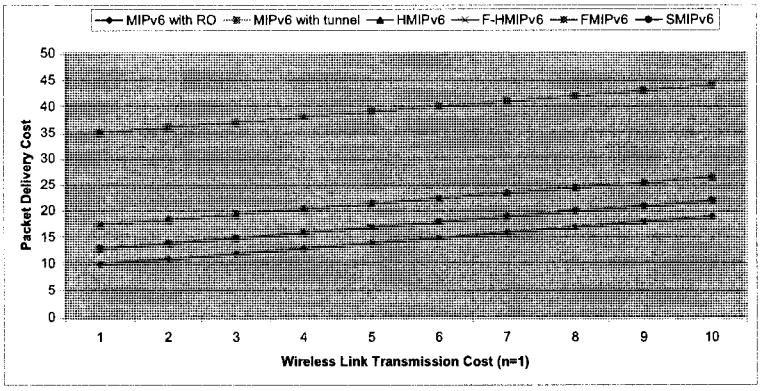

Figure 4.37 Packet Delivery Costs vs. Wireless Link Costs ($N = 1$)

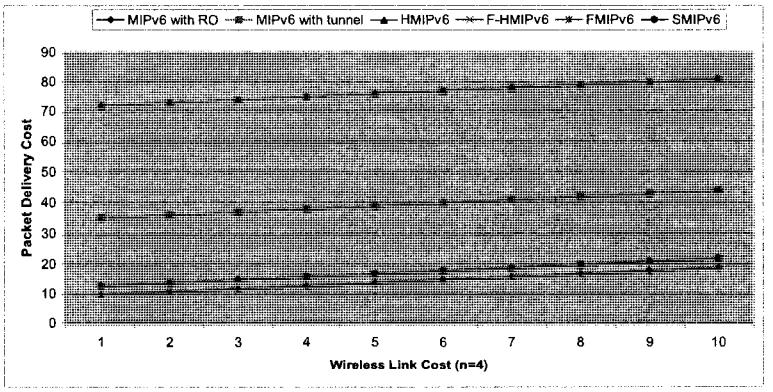

Figure 4.38 Packet Delivery Costs vs. Wireless Link Costs ($N = 4$)

Packet delivery costs versus wireless link costs

Figure 4.37 shows the relationship between the packet delivery costs and wireless link costs for a MAP domain with 1 ring. We observe that the packet delivery costs increase linearly with the wireless link transmission costs. And MIPv6 with tunnel mode requires the highest costs. The mean costs are 39.50 for MIPv6 with tunnel mode, 21.86 for F-HMIPv6 and HMIPv6, 17.30 for FMIPv6 and SMIPv6, 14.30 for MIPv6 with RO mode. Generally, the higher the wireless link transmission cost, the higher the packet delivery costs.

Figure 4.38 shows the relationship between the packet delivery costs and wireless

link costs for a MAP domain with 4 ring. We observe that the packet delivery costs increase linearly with the wireless link transmission costs. And F-HMIPv6 and HMIPv6 require the highest costs. The mean costs are 76.49 for F-HMIPv6 and HMIPv6, 39.50 for MIPv6 with tunnel mode, 17.30 for FMIPv6 and SMIPv6, 14.30 for MIPv6 with RO mode. Generally, the higher the wireless link transmission cost, the higher the packet delivery costs. In addition, increasing the domain size results in higher packet delivery costs for HMIPv6 and F-HMIPv6. We explain this as increasing the number of ARs in a domain causes higher MAP processing costs. Furthermore, HMIPv6 and F-HMIPv6 deliver identical performance, as do FMIPv6 and SMIPv6.

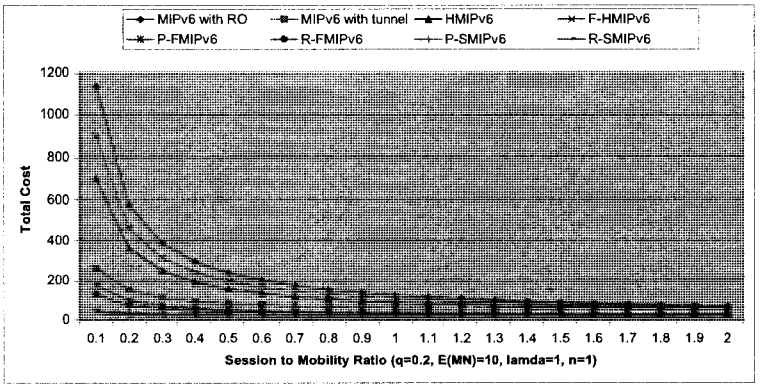

Figure 4.39 Total Costs vs. Session-to-Mobility Ratio ($N = 1$)

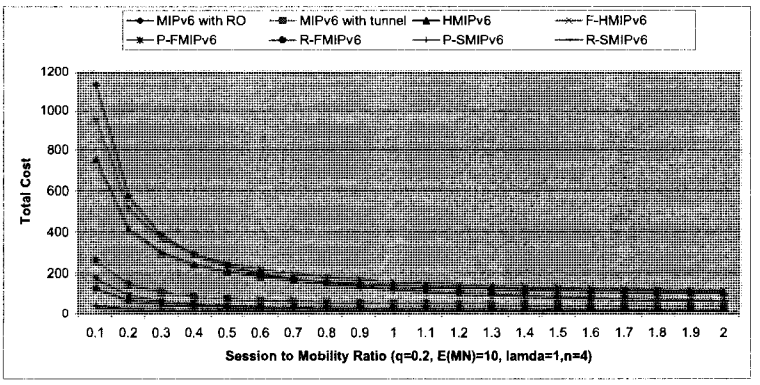

Figure 4.40 Total Costs vs. Session-to-Mobility Ratio ($N = 4$)

Total costs versus session-to-mobility ratio using the random-walk model

Figure 4.39 shows the relationship between total cost and Session-to-Mobility Ratio (SMR) for an MAP domain with 1 ring using the random-walk model. The SMR is defined as the ratio of the session arrival rate over the user mobility ratio, analogous to the Call-to-Mobility ratio (CMR) used to analyze performance in cellular networks. Under the random-walk mobility model, the SMR is defined as $\lambda_s \times E(T)$. The higher the SMR, the lower the total costs. When $SMR \leq 1$, i.e. the session arrival rate is lower than the user mobility rate, signalling costs are more dominant than packet delivery costs over the total costs. As a result, MIPv6 with RO mode requires the highest costs of all schemes. When $SMR \leq 1$, the mean costs are 338.84 for MIPv6 with RO mode, 276.11 for F-HMIPv6, 217.53 for HMIPv6, 101.61 for MIPv6 with tunnel mode, 60.66 for Predictive FMIPv6 (P-FMIPv6), 46.60 for Reactive FMIPv6 (R-FMIPv6), 23.17 for Predictive SMIPv6 (P-SMIPv6) and 18.49 for Reactive SMIPv6 (R-SMIPv6). In addition, as $SMR \geq 1$, the impact of packet delivery costs over total costs becomes much significant. The higher the SMR, the more important the packet delivery costs. In this case, the mean cost is 85.70 for MIPv6 with RO mode, 77.21 for F-HMIPv6, 63.84 for HMIPv6, 50.98 for MIPv6 with tunnel mode, 24.50 for Predictive FMIPv6 (P-FMIPv6), 21.29 for Reactive FMIPv6 (R-FMIPv6), 15.94 for Predictive SMIPv6 (P-SMIPv6) and 14.87 for Reactive SMIPv6 (R-SMIPv6).

Figure 4.40 shows the relationship between total costs and SMR for an MAP domain with 4 ring using the random-walk model. The higher the SMR, the lower the total costs. When $SMR \leq 1$, i.e. the session arrival rate is lower than the user mobility rate, signalling costs are more dominant than packet delivery costs over the total costs. As a result, MIPv6 with RO mode requires the highest costs of all schemes. When $SMR \leq 1$, the mean costs are 338.84 for MIPv6 with RO mode, 330.74 for F-HMIPv6, 272.16 for HMIPv6, 101.61 for MIPv6 with tunnel mode, 60.66 for Predictive FMIPv6 (P-FMIPv6), 46.60 for Reactive FMIPv6 (R-FMIPv6), 23.17 for Predictive SMIPv6 (P-SMIPv6) and 18.49 for Reactive SMIPv6 (R-SMIPv6). Furthermore, as $SMR \geq 1$, the impact of packet delivery costs over total costs becomes much significant. The higher the SMR, the more important the packet delivery costs. In this case, the mean cost is 131.84 for F-HMIPv6, 118.46 for HMIPv6, 85.70 for MIPv6 with RO mode, 50.98 for MIPv6 with tunnel mode, 24.50 for Predictive FMIPv6 (P-FMIPv6), 21.29 for Reactive FMIPv6 (R-FMIPv6), 15.94

for Predictive SMIPv6 (P-SMIPv6) and 14.87 for Reactive SMIPv6 (R-SMIPv6). In addition, increasing domain size leads to the augmentation of total costs for both HMIPv6 and F-HMIPv6. We explain this by the fact that the processing cost at the MAP, especially the routing costs increase due to the augmentation of the number of ARs within a domain. When $SMR \geq 0.4$, F-HMIPv6 requires more costs than MIPv6 with RO mode. When $SMR \geq 0.8$, HMIPv6 requires more costs than MIPv6 with RO mode. However, SMIPv6 always delivers the best performance than other protocols.

Total costs versus session-to-mobility ratio using the fluid-flow model

Figure 4.41 shows total costs and SMRs for a MAP domain with 1 ring using the fluid-flow model. The SMR is defined as λ_s/R_c. As we fix the value of user velocity, the cell crossing rate is fixed. Hence, increasing SMR implies the augmentation of the session arrival rate. In this case, the higher the SMR, the higher the total costs. From the figure, we observe that MIPv6 with RO mode requires the highest costs of all schemes. The mean cost is 67.95 for MIPv6 with RO mode, 33.17 for MIPv6 with tunnel mode, 26.30 for F-HMIPv6, 19.27 for HMIPv6, 16.80 for Predictive FMIPv6 (P-FMIPv6), 14.15 for Reactive FMIPv6 (R-FMIPv6), 9.75 for Predictive SMIPv6 (P-SMIPv6) and 8.86 for Reactive SMIPv6 (R-SMIPv6). When $SMR \geq 0.4$, MIPv6 with tunnel mode requires more costs than F-HMIPv6. Additionally, the proposed SMIPv6 schemes provide better performance than other mobility management solutions.

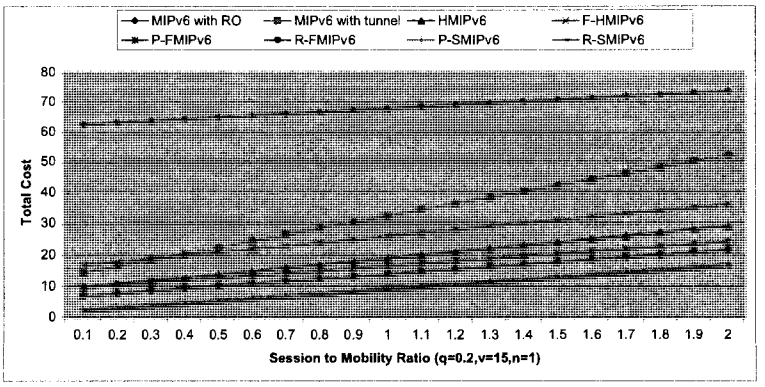

Figure 4.41 Total Costs vs. Session-to-Mobility Ratio ($N = 1$)

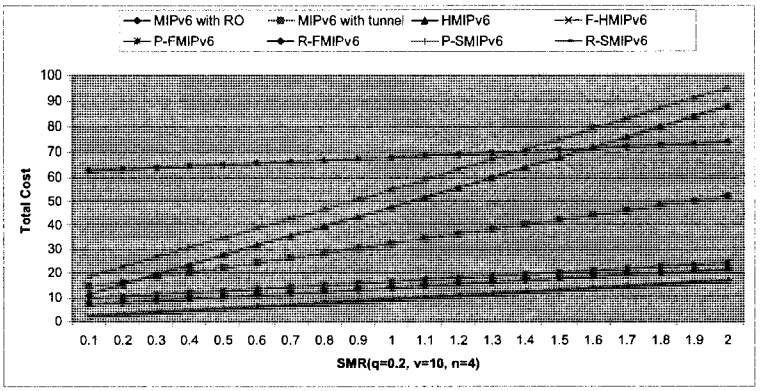

Figure 4.42 Total Costs vs. Session-to-Mobility Ratio ($N = 4$)

Figure 4.42 shows total costs and SMRs for a MAP domain with 4 ring using the fluid-flow model. The SMR is defined as λ_s/R_c. As we fix the value of user velocity, the cell crossing rate is fixed. Hence, increasing SMR implies the augmentation of the session arrival rate. In this case, the higher the SMR, the higher the total costs. From the figure, we observe that MIPv6 with RO mode requires the highest costs of all schemes. The mean cost is 67.95 for MIPv6 with RO mode, 56.75 for F-HMIPv6, 49.52 for HMIPv6, 33.17 for MIPv6 with tunnel mode, 16.80 for Predictive FMIPv6 (P-FMIPv6), 14.15 for Reactive FMIPv6 (R-FMIPv6), 9.75 for Predictive SMIPv6 (P-SMIPv6) and 8.86 for Reactive SMIPv6 (R-SMIPv6). When $SMR \geq 1.4$, F-HMIPv6 requires more costs than MIPv6 with RO mode. When $SMR \geq 1.6$, HMIPv6 requires more costs than MIPv6 with RO mode. When $SMR \geq 0.3$, HMIPv6 requires more costs than MIPv6 with tunnel mode. The increasing of domain size leads to the augmentation of total costs for HMIPv6 and F-HMIPv6. Yet, it has no impact on the costs of MIPv6, FMIPv6 and SMIPv6. We explain this by the fact that the processing cost at the MAP increases due to the augmentation of the number of ARs within an MAP domain. In addition, we find that SMIPv6 always provides better performance than other protocols.

4.7 Simulations

In order to attest the efficiency of our solution, we realized a partial performance of evaluation, which was in relation to the predictive seamless handoff process. In this

section, following the implementation details, the simulation experiments are given. The results are after presented and briefly discussed. The experiments were carried out on OPNET Modeler version 12.0. This simulator was chosen because it is open source and quite renowned in the scientific community. But adaptations to the code and the simulation environment were mandatory to facilitate the implementation.

To evaluate the performance using simulations, we first build a network scenario using campus WLAN (802.11b/g), which has about $1000m$ radio coverage. The advantage of this selection is to have different types of configurations depending on the requirements from system administrator. The network comprises three WLANs, in which three APs are deployed. The network in which locates the AP0 emulates the MN's home network, thus AP0 functions as the Home Agent of the MN. Note that each AP integrates the functionalities of one AR. The MN starts registration at home network, then leaves the coverage area of the AP0, and continues moving into AP1 radio range, then goes to the AP2 coverage area. To implement SMIPv6, AP1 works as the PAR and AP2 functions as the NAR. The line in pink shows the MN's trajectory, shown in Figure 4.41. The background traffic is represented by four CNs: TELNET_CN, EMAIL_CN, FTP_CN and HTTP_CN sending packets to the roaming MN. The topology used for performance analysis is shown in Figure 4.43.

4.7.1 Implementation details

To implement our proposed SMIPv6 schemes, the methodology of implementation is to utilize an existing functional MIPv6 scenario and change the source codes of MIPv6 processes as little as possible. We do not modify any process of MIPv6 but create new ones if necessary.

Figure 4.44 shows the implementation architecture using OPNET Modeler. At the network layer, we define an SMIPv6 network scenario, called *smipv6_scen1.nt*. At the node layer, we define three types of nodes: MN, CN and AR, which are represented by *smipv6_wlan_ethernet_wkstn.nd*, *smipv6_ppp_wkstn.nd* and *smipv6_wlan_ethernet_slip.nd*. At the process layer, we create new libraries for source codes in order to define data structures, which are used by the process during implementing the protocol SMIPv6. We define *ip_dispatch_smipv6_support.pr*, *smipv6_mn.pr*, *smipv6_mgr.pr*, *ipv6_ra_host_smipv6_support.pr* and *ipv6_ra_gtwy_smipv6_support.pr*. Their functionalities are listed as follows:

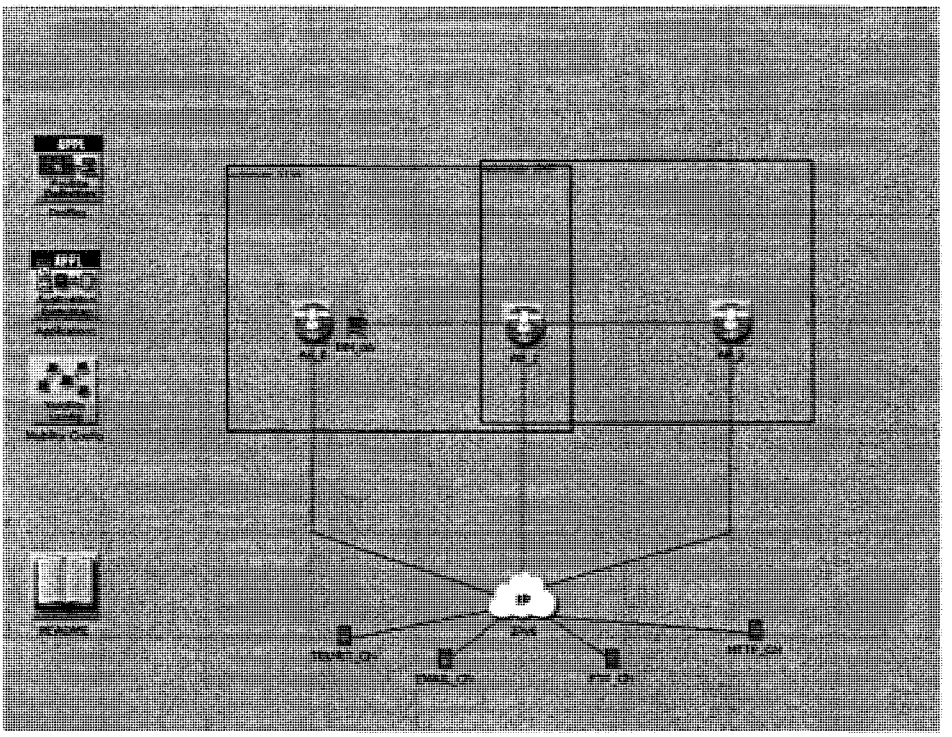

Figure 4.43 Network Topology Used for Simulation

- *ip_dispatch_smipv6_support.pr* is a process which is in charge of packet reception, packet treatment and packet transmission to the corresponding outgoing port according to the routing table.

- *ipv6_ra_gtwy_smipv6_support.pr* is a process that allows sending periodically Router Advertisement (RA) message. Such message is modified with the addition of Basic Service Set Identifier (BSSID). This enables AR to integrate the functionalities of the AP.

- *ipv6_ra_host_smipv6_support.pr* is a process to handle the received Router Advertisement (RA) messages, extract the BSSID of the AP, and lance the process of *smipv6_mgr*.

- *smipv6_mgr.pr* is a process that inherits the process of *ip_dispatch_smipv6_support* and controls the functionality of the protocol SMIPv6. In case where an AR implements such process, the AR handles the mobility related messages, such as

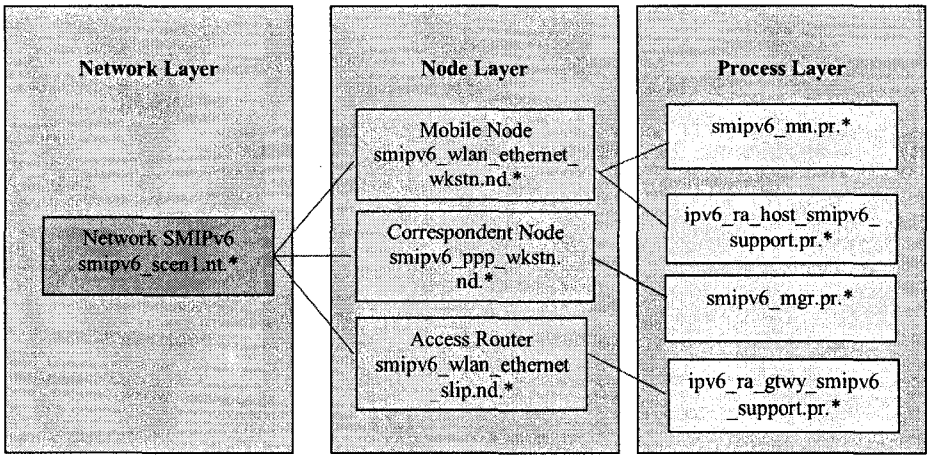

Figure 4.44 Implementation Architecture under OPNET

SBU, RA and SNA. The AR reads the value of attributes, initiates and registers the information about statistics, and displaces the configuration of a node, etc.

- *smipv6_mn.pr* is a process that inherits the process of *smipv6_mgr* and enables MNs to send out mobility related messages, handle the received RA messages, manage some defined timer with respect to the mobility management, etc.

Figure 4.45 shows the implementation process architecture under OPNET Modeler v.12.0.

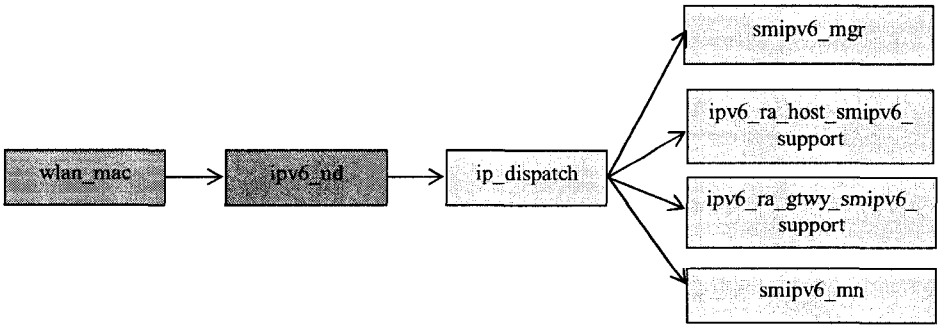

Figure 4.45 Implementation Process Architecture

4.7.2 Simulations results

The simulation experiments gave different results for our proposed SMIPv6 schemes and the OPNET solution for MIPv6. Hereafter, we present the results for the http traffic sent during mobility management, the end-to-end delays during handoff, packet losses caused by the handoff management, and control traffic sent during handoff. Although we show those on the downlink way, i.e. the MN functions as a client while the active CNs act as servers. our discussion applies to the uplink traffic.

HTTP traffic sent during handoff

Figure 4.46 shows the evolution of an HTTP connection between the MN and the CN. Here the MN functions as a HTTP client while the CN acts as an HTTP server. The first part of the figure illustrates the case for Predictive SMIPv6 while the following part indicates the case for MIPv6. We observe that without SMIPv6, the connection is cut off as soon as the MN leaves its home network. From the figure, we find that SMIPv6 deliver better performance than MIPv6.

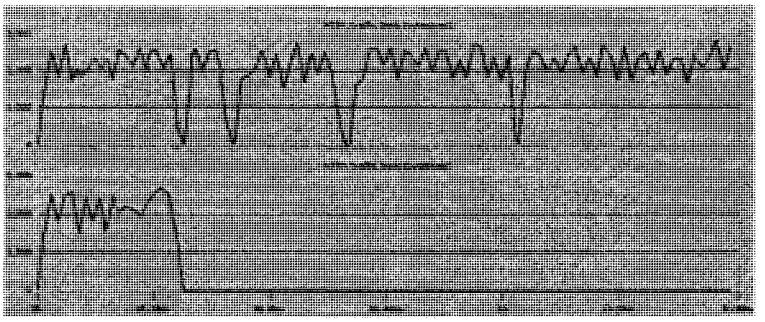

Figure 4.46 HTTP Traffic Sent Comparison During Handoff

End-to-end delays during handoff

Figure 4.47 illustrates the variation of the average WLAN end-to-end delays during handoff. The red shows the performance of MIPv6 while the blue curve shows that of the Predictive SMIPv6. In the MN's home network, the end-to-end delays are similar, on average. The average delay for SMIPv6 is about $5.5ms$ while $9ms$ for

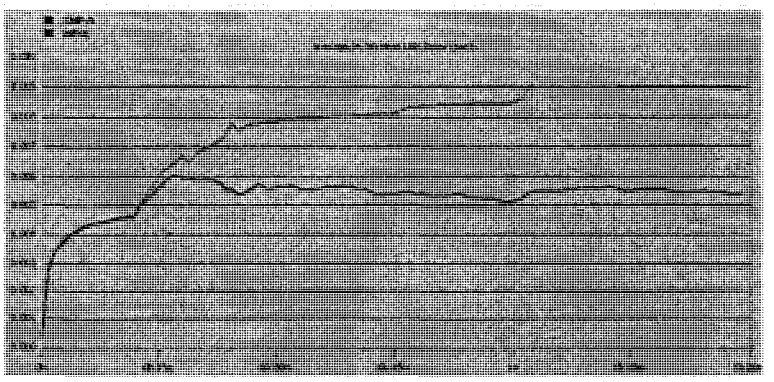

Figure 4.47 WLAN End-to-end Delay Comparison

MIPv6 with RO mode. Hence, predictive SMIPv6 delivers better performance than MIPv6 in terms of end-to-end WLAN delays.

Handoff caused packet drop rate

Figure 4.48 shows the packets drop rate during handoff for predictive SMIPv6 while Figure 4.49 illustrates the packet drop rate for MIPv6. From the figure, we observe that the handoff process in SMIPv6 is much shorter than MIPv6, thus incurs less packet drops.

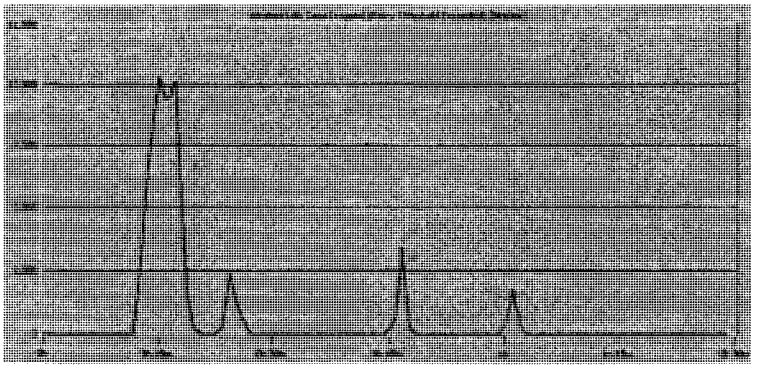

Figure 4.48 SMIPv6 WLAN Data Dropped During Handoff

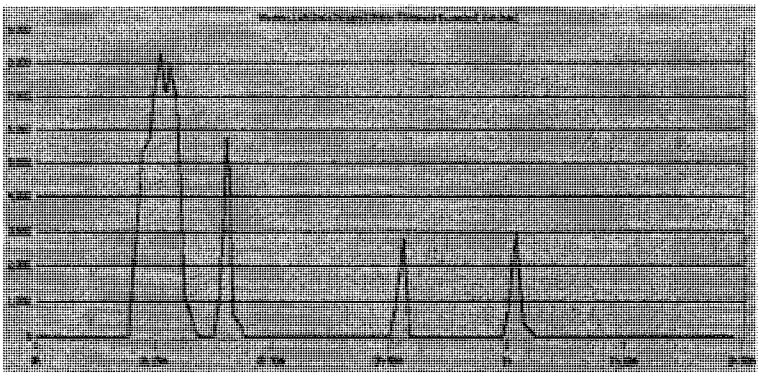

Figure 4.49 Mobile IPv6 WLAN Data Dropped During Handoff

Control traffic sent during handoff

Figures 4.50 and 4.51 show the volume of control traffic sent by the MN during handoff. Generally, control traffic consists of signalling messages exchanged between the MN and other network entities. Sometimes, MIPv6 is better than SMIPv6. But in general SMIPv6 provides better performance than MIPv6, because registration with the PAR is done locally and simply during handoff, thus results in the reduced signalling overhead.

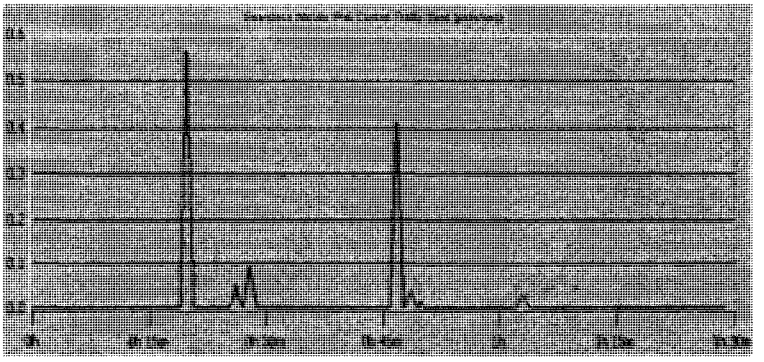

Figure 4.50 SMIPv6 Control Traffic Sent During Handoff

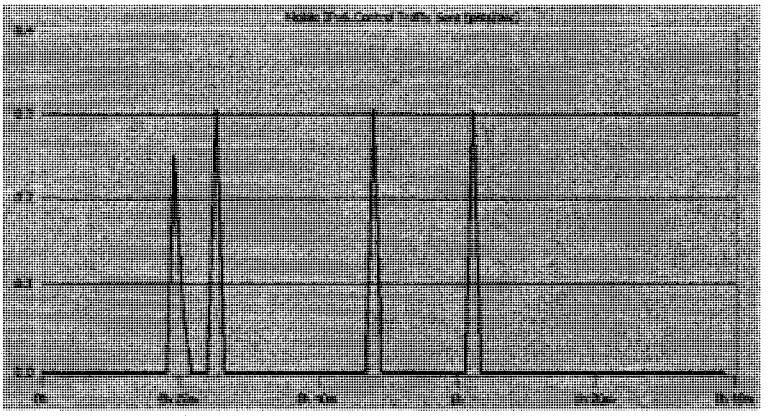

Figure 4.51 Mobile IPv6 Control Traffic Sent During Handoff

CHAPTER 5

PROPOSED FAST MAC LAYER HANDOFF SCHEME FOR MIPV6/WLANs

IEEE 802.11-based wireless local area networks (WLANs) have experienced rapid growth for some years. Increasingly ubiquitous, they are deployed in hotspots found in airports, campuses, malls, etc. to facilitate Internet access for mobile users. In the meantime, Internet service providers (ISPs) can be endowed with substantial productivity increases by enabling moving users to access data in wireless systems. Such facts render handoff management one of the most critical issues for WLANs. However, the IEEE 802.11 specification legacy does not provide handoff support that is sufficiently fast for mobile hosts as they move from one access point (AP) to another. As a result, a number of fast handoff schemes have been proposed in the literature. In this chapter, these fast handoff methods are reviewed and their strengths and weaknesses are exposed. Subsequently, some important design considerations for handoff management in future IEEE 802.11 networks are identified. Moreover, open research issues are highlighted relative to the enhancement of handoff performance and real-time applications support while mobile hosts roam between WLANs. Following the open research issues, we propose new fast medium access control (MAC) layer handoff management schemes for an MIPv6/WLAN environment. The new handoff management scheme aims to support ongoing real-time applications while MNs change their network point of attachment. It consists of minimizing the total number of scanned channels, as well as the probe-waiting time for each examined channel. Performance is evaluated through simulations whose results show that our proposal delivers better performance, compared to the IEEE 802.11b standard, the standard IEEE 802.11b with *MinChannelTime* and two other well-documented solutions in the literature: Selective scanning plus AP Caching and Neighbor Graphs.

5.1 Introduction

The explosion of lightweight hand-held devices with built-in wireless network cards and the significant benefits of ubiquitous Internet access have driven the deployment of wireless local area networks (WLANs). Based on the IEEE 802.11 standard series, WLANs offer users an array of benefits such as user-friendly operations, low cost, large bandwidth, high throughput, etc. so that many multimedia application such as voice over IP (VoIP), media-streaming services, etc. tend to run on top of WLANs. However, when a mobile station moves outside the radio range of its current access point (AP), handoff takes place to ensure the transfer of ongoing calls or data sessions. This handoff procedure involves a series of message exchanges between a mobile station and APs that results in unacceptable delays and packet lost. Fast MAC layer handoff schemes are thus required for IEEE 802.11 WLANs.

This chapter introduces state-of-the-art concepts to improve MAC layer handoff performance. Firstly, the components of the MAC layer handoff procedure in WLANs are introduced to demonstrate the necessity to enhance performance. Then, current approaches to reduce handoff latencies and packet loss rates are outlined in detail, along with their strengths and weaknesses. Open research issues pertaining to handoff management are exposed. And then, potential solutions are proposed to address such challenges. Following the open research issue, we propose new fast MAC layer handoff management scheme for mobile users roaming in an MIPv6/WLAN environment. Performance evaluation is carried out through simulations. We also present some simulation results with detailed performance analysis.

5.2 Background and related work

5.2.1 The IEEE 802.11 handoff process

The IEEE 802.11 Standard defines two operation modes: *infrastructure* and *ad hoc*. In the *infrastructure* mode, an AP comprises a basic service set (BSS) and provides network connectivity to its associated mobile stations. One or more APs comprise an extended service set (ESS) to cover a larger area. In the *ad hoc* mode, two or more mobile stations form a peer-to-peer wireless network without deploying any APs. Note that this chapter focuses only on the infrastructure mode.

An ideal WLAN can provide successive radio signal coverage for mobile stations

in its service area. A mobile station may decide to handoff from one AP to another for mobility reasons, AP load balancing state or signal fading. The legacy MAC layer (L2) handoff process specified in the IEEE 802.11 Standard [108] comprises three phases: *scanning*, *authentication* and *reassociation*. The latter two sub-processes are also referred as re-authentication. The following subsections investigate these three sub-processes in detail.

Scanning

Mobile stations can operate either in a *passive* or *active scanning* mode depending on their configuration parameters.

During *passive scanning*, mobile stations listen for periodic beacon frames generated by APs which announce their presence on each channel and wait for at least a full beacon interval to ascertain beacons receipt from as many APs as possible and at most a *ChannelTime*, on each channel. While scanning, mobile stations cannot transmit frames but they do rather listen for Beacon frames on each channel.

Active scanning involves generating probe request frames and the subsequent processing of received probe response frames. For each channel to be scanned, a mobile station uses active scanning to perform the following [108]:

1. Wait until the *ProbeDelay* time has expired or reception of an indication from the physical layer;

2. Perform wireless medium control using any normal channel access procedure, i.e. carrier sense multiple access with collision avoidance (CSMA/CA);

3. Broadcast a *probe request* frame with the broadcast destination, service set identifier (SSID) and broadcast basic service set identifier (BSSID);

4. Clear and start a *ProbeTimer*;

5. Clear network allocation vector (NAV) and scan the next channel if the wireless medium is idle before the ProbeTimer reaches *MinChannelTime*; otherwise, continue accepting *probe responses* sent periodically by APs within radio range until *MaxChannelTime* and process all received *probe responses*;

6. Clear NAV and scan the next channel.

As indicated above, the passive scanning delay is determined by the number of scanned channels, *ChannelTime* and beacon interval. The probe delay bound, T_p, can be expressed as the following:

$$N \times BeaconInterval \leq T_p \leq N \times ChannelTime \tag{5.1}$$

Where *ChannelTime* refers to the maximal time during which a mobile station listens on each channel and N represents the number of channels available (N equals 32 for 802.11a [109], 11 for 802.11b [110] and 802.11g [111]).

In the same vein, the active scanning delay depends on the number of probed channels, *MinChannelTime* and *MaxChannelTime*. The probe delay bound, T_a, can be expressed as follows:

$$N \times MinChannelTime \leq T_a \leq N \times MaxChannelTime \tag{5.2}$$

Where N depicts the number of channels available; *MinChannelTime* shows the minimal and *MaxChannelTime* shows the maximal probe-waiting time on each channel.

Generally, scanning ends with a set of potential BSSs. Furthermore, passive scanning delays are much longer than those generated by active scanning, since mobile stations are mandated to iterate on all available channels for beacons from APs in range at a set rate (default: $100ms$ per beacon). In addition, mobile stations must dwell on each channel for at least a beacon interval in order to discover as many APs as possible.

Authentication

Authentication aims to identify a mobile station to become a member of a specific BSS, as well as to authorize this mobile station to communicate with other stations in the same BSS. Authentication occurs after a target AP is found. Two authentication methods have been specified for the IEEE 802.11 standard: *open system* and *shared key authentication* [108]. *Open system authentication* involves a pair of frames, an *authentication request* as well as an *authentication response*, which are exchanged between a mobile station and the target AP. Generally, all mobile stations can be authenticated.

Shared key authentication is an optional four-step process that uses the wired

equivalent privacy (WEP) key. A mobile station launches the authentication process by transmitting an *authentication request* to the target AP. Upon receiving this request, the target AP generates a challenge text using a WEP key, and sends an *authentication response* with this challenge text to the mobile station. The latter then encrypts the received challenge text with a shared WEP key and returns an *authentication request* along with the encrypted challenge text to the target AP. Afterwards, the target AP decrypts this request with the shared key and compares the original challenge text with the decrypted one. When they are identical, the target AP transmits an *authentication response* which confirms a successful authentication.

Regardless of the authentication method used, the IEEE 802.11 Standard requires mutually acceptable, successful authentication. Furthermore, authentication is required before an association can be established. Due to current security flaws in open system and shared key authentication, the authentication methods specified in IEEE 802.11 are superseded by IEEE 802.11i [112]. However, considering compatibility, IEEE 802.11i allows open system authentication and exchanges authentication messages after the reassociation phase [112] [113].

Reassociation and association

Association consists of establishing AP and mobile station mapping and enabling station invocation of the distribution system services whereas *reassociation* enables an established association to be transferred from one AP to another [108].

Reassociation is an important component of L2 handoff after a successful authentication. Since the IEEE 802.11 Standard specifies that each mobile station must be associated with a single AP at any given time [108] and a mobile station must issue a *reassociation request* to the new AP during handoff. This request frame contains the previous BSSID, the mobile station's MAC address, etc. and triggers the inter-access point protocol (IAPP) [114] to deliver relevant context information.

Upon receiving this frame, the new AP sends an *access-request* message to the RADIUS (remote authentication dial-in user service [115]) server, which then looks up the IP address of the previous AP and verifies the BSSID, before returning an *access-accept* message to the new AP. This message contains the previous AP's IP address along with security block items required to establish a secure communication channel between APs. After exchanging security elements through *send-security-block* and *ACK-security-block* packets, both APs have obtained sufficient information to

encrypt all further packets. Afterwards, the new AP sends an encrypted *move-notify* packet to the prior AP asking for the context of the concerned itinerant station. Upon verifying the mobile station's association, the old AP removes the mobile station from its association table and replies an encrypted *move-response* packet to the new AP, including the concerned *context block*. Then, the new AP adds the mobile station into its association table and broadcasts a *layer 2 update* frame to inform any layer 2 devices, such as bridges and switches, so that they can update their forwarding table for the specific mobile station. At last, the new AP sends a *reassociation response* to the mobile station [114], [116], [117]. The overall handoff process is completed when this response is received. Figure 5.1 illustrates the overall MAC layer handoff process for WLANs.

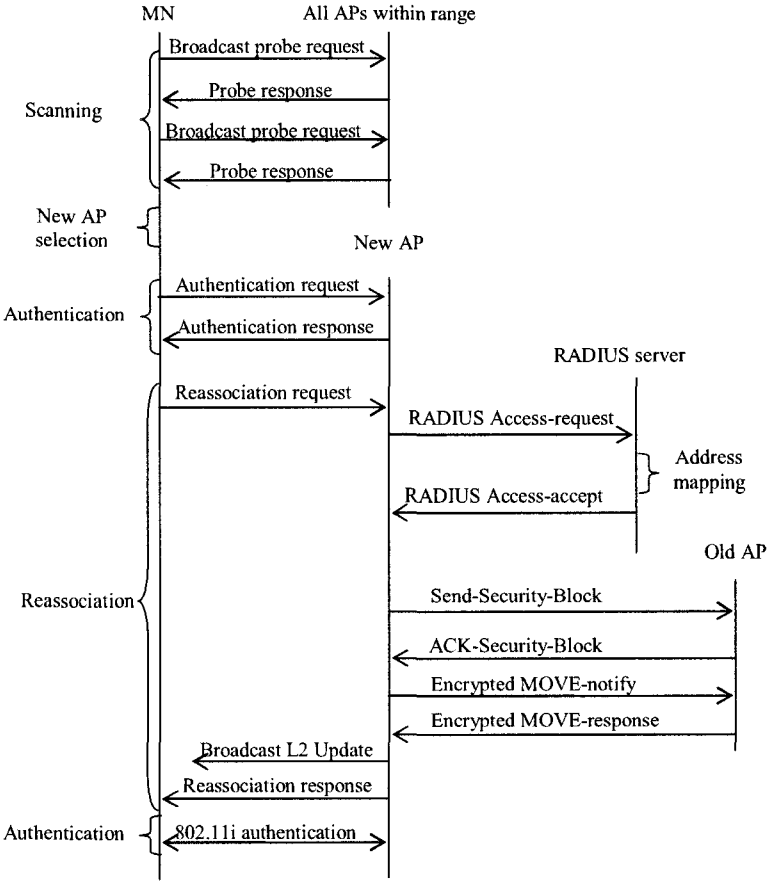

Figure 5.1 MAC Layer Handoff Process in WLANs

Briefly, the inter-access point protocol (IAPP) allows an AP to communicate with other APs in a common ESS, while minimizing opportunities for the transmission of mobile stations'security information over the air. However, context transfer using IAPP results in additional delays during handoff.

Numerous studies have been conducted in order to improve MAC layer handoff performance in terms of handoff delays (time required to complete scanning, authentication and reassociation) and packet loss rates for mobile hosts roaming in IEEE 802.11 networks. This chapter builds on previous work in [113] [118], focusing on the investigation of typical fast handoff schemes in recent literature. To simplify the analysis, two categories are investigated, namely those that reduce delays pertaining to probing and re-authenticating.

5.2.2 Fast handoff schemes to reduce probe delays

Probe delays consist of the main contributor to the overall MAC layer handoff latency [119] and most recently proposed handoff schemes aim to reduce this lengthy delay. These schemes can be further classified into: fast scanning, bypass scanning and cross-layer design.

Fast scanning

Fast scanning methods rely on reducing the number of probed channels, the time taken on each channel, scanning-related timers, such as *MinChannelTime* and *MaxChannelTime* for active scanning, *ChannelTime* and *beacon interval* for passive scanning, etc. Such methods can be further classified into full and selective scanning.

Full scanning means that all available channels are probed while the values of *MinChannelTime*, *MaxChannelTime*, probe-waiting time and/or *beacon interval* are optimized. Usually, full scanning is based on the assumption that mobile stations have no pre-knowledge of existing neighboring APs within their range when the handoff occurs and, as a result, all available channels must be searched consecutively. There are several full scanning methods. Here are some examples:

- *The tuning technique* [120] aims to find an optimal value for *MinChannelTime* and *MaxChannelTime* in order to reduce active scanning delays. Through rigorous calculations, the authors conclude that *MinChannelTime* and *MaxChan-*

nelTime should be set to $1.024ms$ and $10.24ms$ respectively. Furthermore, link-layer handoff detection delays can be reduced upon the loss of three consecutive frames.

- *Intelligent channel scanning* [121] is designed to minimize the probe-waiting time on each channel. Instead of waiting for $MaxChannelTime$ on a busy channel, mobile stations stop searching immediately once they have collected all available probe responses. Continuous monitoring techniques are thus used by mobile stations to evaluate the number of APs (n) on a probed channel. Accordingly, mobile stations stop scanning after collecting a maximum of n responses on a specific channel.

- *SyncScan* [122] replaces the active scanning procedure with passive channel monitoring on nearby APs. Furthermore, a continuous tracking technique is devised by synchronizing short listening periods at mobile stations with regulated periodic beacon transmission from APs. As a result, mobile stations can passively scan by switching channels at the exact moment a beacon is about to arrive. To do so, a staggered periodic schedule of beacon periods is created and spread across channels. For example, all APs operating on Channel 1 are forced to broadcast beacons at time T , APs on Channel 2 broadcast beacons at time $(T + d)$, APs on Channel 3 broadcast beacons at time $(T + 2d)$, and so on. Therefore, if a mobile station connected to an AP on Channel c receives beacons from Channel c at time T_c, it can receive beacons from APs on Channel $(c + 1)$ at time $(T_c + d)$.

  In a nutshell, SyncScan reduces the cost of continuous scanning and yields better handoff decisions. However, time synchronization is a critical issue amongst all neighboring APs. On the other hand, multiple APs operating on the same channel attempt to generate and broadcast beacons simultaneously, hence bringing about beacon conflicts. Consequently, more collisions take place on wireless mediums. This side effect reduces productivity on wireless links and system throughput. In addition, more packets are lost while mobile stations explore other channels.

- *Smooth handoff schemes* [123] focus on smooth channel scanning by classifying channels into different groups. Rather than scanning all channels consecutively,

once a group of channels is scanned, the mobile station pauses before switching back to normal data transmission mode. This scheme works as a scheduled full scan.

Selective scanning : Instead of probing all available channels individually, *selective scanning* reduces the number of channels required to discover APs. Thus, probe delays are significantly minimized compared to the full scanning method. A number of selective scanning approaches have been proposed in the literature. Here are some examples:

- *The channel mask schemes* [124] allow mobile stations to selectively scan channels with a mask built in previous scans.
- *The neighbor graph (NG) methods* [125] [126] allow mobile stations to scan only channels which neighbor APs operate on, which drastically reduces the number of probed channels. Together with the NG method, several schemes are proposed to minimize probe-waiting time on each neighboring channel, i.e. NG-pruning approach [125], unicast probe request with fast switching between each probed channel [126]. A modified NG solution [126] is developed with a unicast probe request to a neighbor AP previously selected by an NG server. And to shorten probe-waiting time, mobile stations switch to another channel when they receive a probe response from the specific AP [126].
- *Handover assisted by geolocation information* [127] aims to predict the next AP and the associated subnet using the mobile station's position and the topology information of domain APs.
- *Sensor network-assisted handoff* [128] is designed to reduce scanning delays by limiting the number of probed channels during handoff. Before the actual handoff, the sensor network is deployed and overlaid with a WLAN. Sensors collect the parameters of neighboring APs. Consequently, prior to a handoff decision, a mobile station broadcasts an *AP list request.* Equipped with this request, sensors located within range act as neighboring relay nodes and reply with an *AP list response* that contains all required information about surrounding APs. Accordingly, at the moment when the handoff occurs, mobile stations solely scan channels indicated on the list they have received.

Bypass scanning

As scanning is the most time-consuming component of the overall link-layer handoff, certain solutions focus on bypassing scanning to eliminate the probe latency. For example, the caching technique or multiple radio interfaces deployed either at the AP or at the mobile station uses to decouple scanning with handoff so that mobile stations can search proactively for alternate APs while being associated with an AP and interleaving data communication. Here are some examples:

- *The caching technique* aims to buffer neighboring APs' information and exploit this information to accelerate the scanning procedure. Usually, this approach implies trivial modifications at the mobile station and the size of caching tables are either fixed [124] or dynamic [129]. AP caching [124], neighbor graph caching (NGC) [130] and adjacent APs [131] represent typical examples.

- *Multiple radio interfaces at mobile station approach* [132]-[135] is a physical layer approach designed to completely eradicate scanning. During handoff, one wireless interface is used for normal data communication with the associated AP, while the other is used to search surrounding APs and find a candidate one to reassociate with. *MultiScan* [132] consists of a relevant example of such a strategy.

- *Multiple radio interfaces at AP approach* [136]-[138] also consists of a physical layer method conceived to eradicate scanning. Additional radio interfaces eavesdrop on neighboring channels to rapidly detect mobile station movements [136] [137]. Moreover, this additional radio transceiver can be used to search neighboring stations located within range and control their handoff operations [138].

- *Pre-scanning methods* allow mobile stations to scan neighboring APs while they are associated with an AP. They interleave scanning with data communication, such as pre-scanning with selective channel mask [129], periodic scan [131], proactive scan with smart triggers [139], pre-active scan [140], anticipated handover [141], anticipated scanning [142], continuous monitoring with smart triggers [143], etc.

- *The location-based fast handoff method* [144] allows mobile stations to select potential APs by predicting the path of their movement. By doing so, a location

server is deployed to provide APs' information to mobile stations so that they can reassociate with the new AP directly, without scanning channels. However, this method relies on precise localization methods.

Cross-layer design

It may be beneficial that mobile stations maintain IP connectivity during their movements: this brings about additional requirements for efficient mobility management in wireless LANs. This new research objective provides support for seamless handoff and real-time multimedia applications in WLANs. Generally, user mobility is managed using MIPv6 [7] or MIPv4 [5]. The typical handoff procedure comprises movement detection, new address configuration and registration, etc. and turns in unacceptable delays for real-time services. Therefore, several handoff schemes have been proposed to improve handoff performance using cross-layer design strategies. Some of them are briefly introduced as follows:

- *Beacon with sufficient IP layer information* [145] [146] allows an enhanced AP to assist and handle fast new address configuration by inserting IP layer information, such as router advertisement [145] and network prefix information [146] into beacon frames. This approach makes it possible to drastically decrease overall handoff latencies (both at MAC layer and IP Layer).

- *IP-IAPP scheme* [147] [148] enhances APs with advanced routing functionalities so that they act as mobility agents for mobile stations. They are also responsible for IP mobility management.

- *Link-layer triggers and topology information-aided fast handoff* [149] use pre-handoff triggers to discover agents or address configuration prior to IP layer handoffs. In addition, post-handoff triggers are applied to eliminate movement detection delays.

In the following sections, we describe three typical fast scanning approaches: selective scanning and AP caching schemes, neighbor graph and NG-pruning schemes, and handoff assisted by geolocation information.

Selective scanning and AP caching schemes

Shin et al. [124] propose channel masks and AP caching schemes to reduce probe delays to a level where voice over IP (VoIP) communication becomes seamless. These schemes focus on reducing the probing time of non-existing channels through selective scanning, as well as the frequency of selective scanning using caching techniques.

Selective scanning is performed using channel masks that are built when drivers are first loaded on mobile stations. A full-scan is conducted through broadcasting probe requests on all available channels. When a probe response arrives, a channel mask is set for the examined channel. In addition, channel masks are set by default for Channels 1, 6 and 11 as they are most likely to be used in well-configured wireless networks. Once all channels are scanned, the mobile station selects the best AP, based on the received signal strength. Then, it performs authentication and reassociation with the newly selected AP. Accordingly, channel masks are updated by removing the currently associated AP from the channel mask. When the ensuing handoff occurs, only channels equipped with a mask are probed. If no APs are found, the channel mask is inverted and a new selective scan is required. If no APs are found, a full scan is conducted to build new channel masks.

AP caching scheme consists of a cache table where the current AP's MAC address is indexed as a key. The list composed of the adjacent APs discovered during the scanning phase corresponds to the key. When a mobile station becomes associated with an AP, the latter is entered as a key into the cache. Cache entries are checked when a handoff is launched. If no entries are found, (this case is called cache miss), the mobile station performs selective scanning and inserts two APs with the highest received signal strength into the cache. If an entry is found in the cache, (this case is called cache hit), the station initially attempts to connect to the first AP. Once reassociation is done with success, the handoff is over; otherwise, it will try to associate with the second AP in the cache. When reassociation with this AP is done successfully, the handoff is over; otherwise, selective scanning is necessary to build new channel masks and find new APs for further reassociations. Figure 5.2 illustrates the selective scanning and AP caching schemes.

The symbol m/n denotes AP_m operating on Channel n. Each mobile station maintains a channel mask built after a full scan. The corresponding channel mask is shown in Table 5.1. During ensuing handoffs, mobile stations only scan mask-wearing channels. In Figure 5.2, a mobile station currently connected to AP_1 only

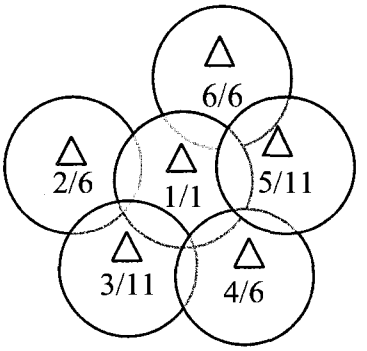

Figure 5.2 An Example for Selective Scanning and AP Caching Schemes

scans Channels 1, 6 and 11 during the handoff. Table 5.2 illustrates the examples of cache table. The cache table 1 allows a mobile station to reassociate with AP_3 and AP_4 without scanning while the cache table 2 results in a cache miss during handoff.

Table 5.1 Channel Mask Table for Figure 5.2

1	2	3	4	5	6	7	8	9	10	11
1	0	0	0	0	1	0	0	0	0	1

Table 5.2 Examples of Cache Table

(a) Cache Table 1

Key	1st AP	2nd AP
1/1	3/11	4/6
2/6	1/1	3/11
6/6	1/1	5/11

(b) Cache Table 2

Key	1st AP	2nd AP
1/1	-	-
2/6	1/1	3/11
6/6	1/1	5/11

In short, the channel masks scheme presents a fast selective scanning technique as mobile stations solely need to scan channels endowed with masks after a full scan whereas the AP caching scheme reflects a bypassing scanning approach. These schemes result in enhanced handoff performance. However, as caching tables are built from previous scanning results, mobile stations are likely to select an incorrect AP during handoff, thus triggering false handovers. Furthermore, cache misses hinder network performance.

Neighbor graph and NG-pruning schemes

Neighbor graph (NG) and NG-pruning Schemes are proposed to enhance MAC layer handoff performance when mobile stations roam in WLANs [125]. This newly discovery method aims to reduce the total number of probed channels as well as the probe-waiting time on each channel. NGs dynamically capture the mobility topology of wireless networks [150] [151] to assist mobile stations in making decisions regarding whether or not a channel needs to be scanned. Meanwhile, using non-overlapping graphs, mobile stations can find out whether to wait longer for probe responses on an examined channel, before the *MaxChannelTime* expires.

Before generating an NG, a mobility graph is defined to aggregate stations' mobility traces in WLANs. Using adaptive estimation techniques, an NG is created by abstracting handoff relationships between adjacent APs. A non-overlapping graph is created by abstracting non-overlapping relationships amongst APs. Three methods are used to implement NGs:

- *The centralized method*: an NG server is used to restore the NGs and provide mobile stations with a NG as they join the network.
- *The distributed method*: each AP stores its local NG and mobile stations retrieve this graph from the AP after reassociation [150] [151].
- *The personal or user-oriented method*: each mobile station keeps track of its mobility patterns to create its own NG.

An NG can be defined as: $G = (V, E)$ where $V = \{AP_i | i = 1, 2, ..., n\}$ (represents the set of all APs of a wireless system), $E = \{(AP_i, AP_j) | i \neq j\}$ and (at least one mobile station handoffs from AP_i to AP_j). A non-overlapping graph can be defined as: $G = (V, E)$ where $V = \{AP_i | i = 1, 2, ..., n\}$ (depicts the set of all APs of a wireless system), $E = \{(AP_i, AP_j) | i \neq j\}$ and (mobile stations cannot communicate with AP_i and AP_j simultaneously with acceptable link quality).

An example of APs location map and the corresponding NG is illustrated in Figure 5.3. The symbol m/n denotes AP_m operating on Channel n. In this figure, mobile stations associated with AP_1 can handoff to AP_2 and AP_5, AP_2 to AP_1 and AP_3 to AP_1 and AP_4. Using the NG scheme, mobile stations covering by AP_1 only scan Channel 6 during the handoff. The number of probed channels is thus drastically reduced using the NG.

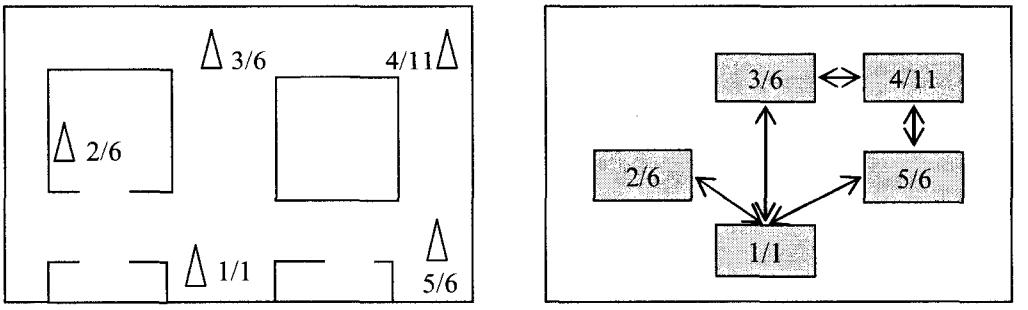

Figure 5.3 APs Location Map and Corresponding NG

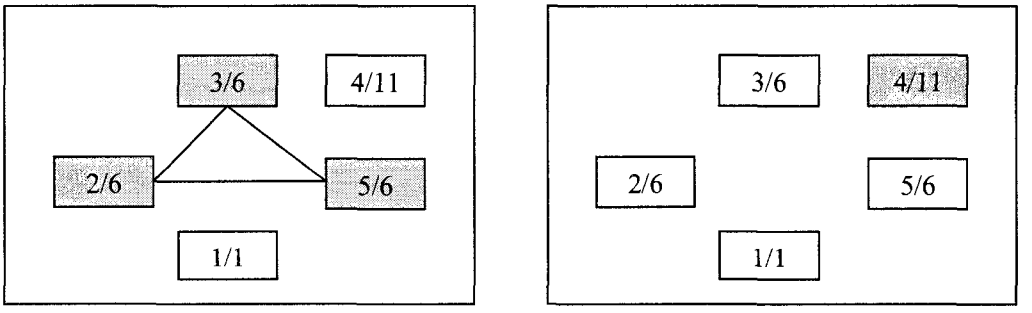

Figure 5.4 Non-overlapping Graphs for Channel 6 (left) and Channel 11 (right)

Non-overlapping graphs for the aforementioned NG are depicted in Figure 5.4. Based on these non-overlapping graphs, the NG-pruning technique makes it possible for a mobile station scanning Channel 6, to stop searching the same channel upon receiving a probe response from AP_2 or AP_3 or AP_5, as they do not overlap. NG-pruning scheme can thus drastically reduce probe-waiting time on each probed channel. For example, mobile stations probing Channel 11 stop scanning when they receive a probe response from AP_4, since a single AP operates on Channel 11.

Briefly, NG and NG-pruning schemes allow mobile stations to scan only a subset of all available channels and spend less waiting time on each probed channel, compared to the conventional handoff process defined in IEEE 802.11. However, mobile stations require global knowledge of the wireless environments before the actual handoff. Moreover, the quality of the NG scheme can be impaired by important topol-

ogy changes in cases where APs are added or removed [125]. Furthermore, building such graphs involves a considerable amount of time and maintaining mobility graphs happens to be a complex task.

Handoff assisted by geolocation information

A fast handoff method using geolocation information provided by a global positioning system (GPS) is proposed to reduce MAC and IP layer handoff latencies [127]. All mobile stations are equipped with a GPS receiver which estimates the station's position and reports the obtained measurements to the station every second.

The station calculates the distance (d_1) between its current and previous locations using a Haversine formula. If ($d_1 > 1$) meter, the mobile station sends a *location update* (LU) message that includes its coordinates to a GPS server. Note that this server keeps a list of domain APs' topology information and their relative parameters, such as $< AP\_ID, (x, y), Channel, SSID, IPv6\_Prefix >$.

Upon receiving the LU message, the GPS server assesses the distance (d_2) between the mobile station and its current AP. If ($d_2 > G$), G denotes a pre-defined threshold, the GPS server selects the AP closest to the mobile station as the new AP and sends a *handover initiate* (HI) message to the mobile station. This message contains the target AP's ID, its operating channel and IPv6 prefix.

After receiving this message, the mobile station launches a handover by sending out a *probe request* over the new AP's channel and waits for a *probe response* from the target AP. When the mobile station receives a *probe response*, it launches authentication and reassociation processes. Simultaneously, the mobile station looks up the new AP's IPv6 prefix. If this prefix differs from the previous one owned by the station, the mobile station starts IP layer handover immediately, without waiting for a *router advertisement* (RA).

Subsequently, the mobile station configures a new IPv6 address and sends a *binding update* (BU) to its home agent (HA), which then replies with a *binding acknowledge* (BAck) message to the mobile station, thus completing the handoff when the BA is received.

In summary, GPS-assisted Handoff scheme consist of a fast, selective scanning approach since mobile stations only need to scan a single channel during the MAC layer (L2) handoff. It also embodies a cross-layer design method as mobile stations can find out their exact new subnet prefix before completing the L2 handoff. Moreover,

it contains a new fast IP layer (L3) movement detection method without the support of *router advertisements*. However, as this handoff scheme uses a GPS server, it introduces a centralized system and the server becomes a traffic flow bottleneck. In addition, those pre-configured parameters have a significant impact on system performance which could be hindered by certain values. Moreover, in cases of fast movements performed by mobile stations, it becomes unreliable for the GPS server to select a new AP as it is likely to choose an inappropriate AP, leading to wrong handoff decisions.

5.2.3 Fast handoff schemes to reduce re-authentication delays

IEEE 802.11 defines *open system* and *shared key authentication* Methods. *Open system authentication* admits all stations in the distribution system while *shared key authentication* relies on wired equivalent privacy (WEP) to demonstrate the knowledge of a WEP encryption key. However, given that numerous security flaws [113] [152] render both methods vulnerable to attacks, they have been replaced by IEEE 802.11i [112], a standard designed to enhance 802.11 security aspects, by introducing *key management* and *establishment* mechanisms, along with *encryption* and *authentication* improvements [152].

IEEE 802.11i incorporates IEEE 802.1X [153] as its authentication enhancements. As IEEE 802.1X is commonly deployed in many IEEE 802 series standards and uses a *remote authentication dial in user service* (RADIUS) [115] server to manage *authentication*, *authorization* and *accounting* (AAA) related activities, re-authentication (including authentication and reassociation) processes result in important delays. Therefore, fast handoff schemes to reduce this lengthy latency have been proposed in the literature. An overview of fast authentication methods is provided in [113], most of which are designed for intra-extended service set (intra-ESS) handoffs. Some typical fast authentication methods are introduced in the following paragraphs.

IEEE 802.11i pre-authentication

Specified in IEEE 802.11i and designed for mobile stations already associated with an AP in the ESS, pre-authentication is launched by mobile stations which act as IEEE 802.1X supplicant [153]. It allows mobile stations to authenticate multiple APs at

once [112], rendering authentication independent from roaming.

The roaming station sends an *EAPOL-start* (extensible authentication protocol over LANs-start) message to the new AP via its associated AP. Then, the new AP acts as authenticator to initiate an IEEE 802.1X authentication process by transmitting an *EAP-request/identity* (extensible authentication protocol request/identity) to the mobile station.

Following that, the mobile station returns an *EAP-response/identity* message to the new AP. Subsequently, the new AP forwards a *RADIUS-access-request* message to the authentication server which replies with a *RADIUS-access-challenge*. Thereafter, the new AP forwards this challenge text to the mobile station in an *EAP-request/auth* message. The mobile station then encrypts the challenge text using a shared secret with the authentication server, and transmits an *EAP-response/auth* to the new AP which forwards a *RADIUS-access-request* containing the encrypted challenge text to the authentication server.

This server decrypts the challenge, which is then compared to the original. When identical, the server returns a *RADIUS-access-accept* message to the new AP, which, in turn, forwards an *EAP-success* message to the mobile station [113], [152]. Upon a successful 802.1X authentication, a shared secret is created and cached between the new AP and the mobile station.

Briefly, pre-authentication allows mobile stations to prevent reassociation during re-authentication, thus leading to significantly shorter delays. However, pre-authentication also introduces new opportunities for denial of service (DoS) attacks and unnecessary burden on the authentication server [113].

Proactive key distribution schemes

Proactive key distribution schemes, including the *proactive neighbor caching* (PNC) [151] and the *selective neighbor caching* schemes [154], are designed to reduce authentication latency by pre-distributing key material one hop ahead of mobile stations' movement. Neighbor graphs are used to dynamically capture the mobility topology of a WLAN with the purpose of pre-distributing mobile stations' contexts to all neighbor APs [151]. The PNC scheme consists of pre-positioning mobile stations' contexts to all neighboring APs while SNC scheme pre-distributes the context to a set of selective APs, based on handoff probabilities.

Figure 5.5 illustrates the proactive neighbor caching scheme. The symbol m/n

denotes AP_m operating on Channel n. When a mobile station associates with AP_2, its context information is propagated to all neighboring APs, $\{AP_1, AP_3, AP_4\}$. When this mobile station handoffs to AP_4, no additional authentication is required since AP_4 has received and cached the mobile station's credentials. Simultaneously, the mobile station's contexts are removed from other non-neighboring APs, i.e. AP_1. When reassociation occurs, the context information is broadcasted to all neighbors of the AP_4.

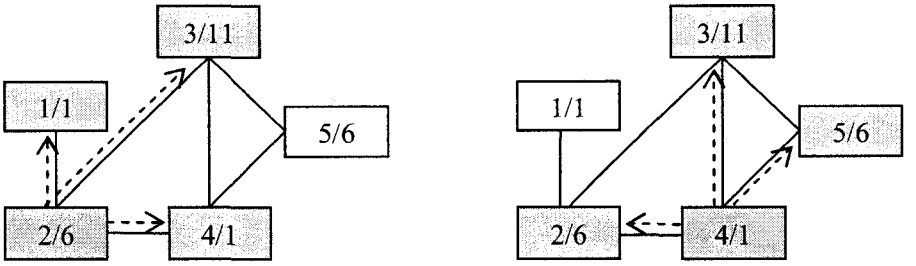

Figure 5.5 The Proactive Neighbor Caching Scheme

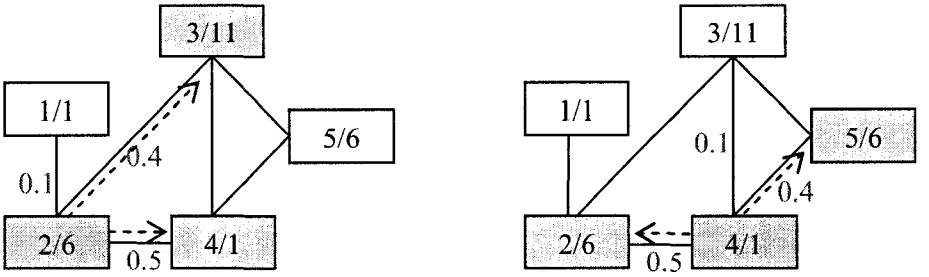

Figure 5.6 The Selective Neighbor Caching Scheme

Figure 5.6 illustrates the selective neighbor caching scheme. The symbol m/n denotes AP_m operating on Channel n. Suppose that the predefined handoff probability threshold is 0.3. The SNC scheme allows mobile stations' contexts to be transferred to surrounding APs, whose handoff probabilities are equal to or greater than 0.3. When a mobile station associates with AP_2, its context information is propagated to a set of selective neighboring APs, $\{AP_3, AP_4\}$, according to handoff probabilities.

As a result, when this mobile station handoffs to AP_4, normal authentication is not required since AP_4 has received and cached the mobile station's credentials. Simultaneously, the mobile station's contexts are removed from other non-neighboring APs. When reassociation occurs, the context information is broadcasted to AP_4's selected neighbors $\{AP_2, AP_5\}$. If this mobile station continues roaming and handoffs to AP_3, it must perform legacy Re-authentication as AP_3 has removed the related context.

In a few words, proactive key distribution schemes can reduce re-authentication delays since mobile stations' contexts are transferred to neighboring APs during (re)association. However, these schemes result in high signaling overhead, especially in WLANs that contain a very dense population of mobile users. To reduce context transfer signaling costs in the PNC scheme, the SNC scheme adds neighbor weights to all edges of the neighbor graph. Neighbors' weight translates into handoff probabilities for each neighboring AP and is generated by monitoring mobile stations' handoff patterns amongst APs. Using the SNC scheme, mobile stations' contexts are solely transferred to neighboring APs with higher handoff probabilities. The SNC scheme provides similar handoff performance when mobile stations handoff to a carefully-selected AP. However, performance degradation is noticed when stations handoff to an AP without cached context. In addition, as handoff probabilities consist of variable factors, the SNC scheme introduces additional complexity compared to the PNC scheme.

The predictive authentication scheme

A predictive authentication scheme is proposed to reduce re-authentication delays using frequent handoff regions (FHRs) [155]-[157]. This scheme is referred to as the FHR scheme in [11]. Mobile stations' authentication information is proactively distributed to multiple APs chosen by an FHR selection algorithm that considers the mobile stations' mobility patterns, service classes, etc. The FHR comprises a subset of adjacent APs most likely to be visited by mobile stations in the near future. The essence of the approach is that when a mobile station enters an AP's covered area, it performs authentication procedures for multiple APs in the FHR. As a result, when it handoffs to an AP in the specific FHR, re-authentication delays are eliminated as the mobile has already registered and authenticated this AP.

Figure 5.7 depicts the operation of the FHR scheme. A mobile station is associated with AP_2 and its corresponding FHR comprises $\{AP_1, AP_2, AP_4\}$. The mobile sta-

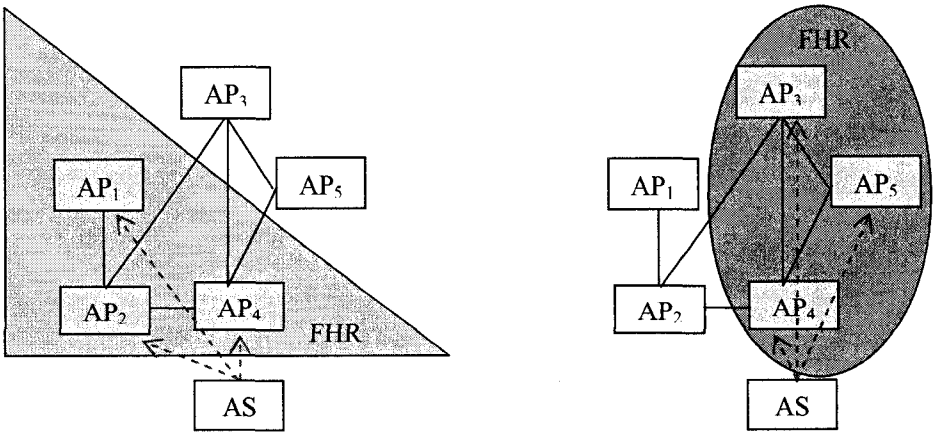

Figure 5.7 The Frequent Handoff Region Scheme

tion's authentication information is propagated to these three APs. As a result, when the mobile handoffs to AP_4, a further authentication procedure is not required, as the Authentication Server (AS) has already dispatched the station's context to AP_4. When connected to AP_4, the mobile station generates a new FHR $\{AP_3, AP_4, AP_5\}$, and its pertaining authentication material is propagated to the APs in the new FHR by the AS.

The FHR scheme is based on mobility predictions to dispatch mobile stations' authentication information to a set of APs in frequent handoff regions. Authentication delays are eliminated in cases where mobile stations handover to an AP within the FHR. However, as the FHR is based on centralized authentication, the authentication server creates bottleneck problems.

5.3 Handoff related open research challenges

Even though a collection of fast handoff schemes has been proposed for mobile stations roaming in WLANs, *seamless mobility support* remains an important and challenging issue. On the other hand, *real-time service support* for roaming users consists of another complex issue. So far, several IEEE working groups have pooled their efforts to improve the conventional IEEE 802.11 Standard: 802.11e strives to enhance quality of service (QoS) [158], 802.11i to reinforce security [112], 802.11f to upgrade the

inter-access point protocol (IAPP) [114], 802.11k [159] to manage radio resources, 802.11r [160] to support fast roaming (transit), 802.21 [63] to improve handoff between heterogeneous wireless networks, etc.

Regarding fast handoff support, the IEEE 802.11r working group is drafting a protocol to facilitate the deployment of IP-based telephony over IEEE 802.11-enabled phones by speeding up handoffs between APs or cells in a WLAN. Moreover, as mobile devices with multiple interfaces emerge [161], a set of *technical management* issues needs to be taken into consideration to provide *seamless connectivity* from one interface to another. In this context, the IEEE 802.21 working group is standardizing a general interface to manage network interface cards as well as extending their work to hide network heterogeneity from end users.

Supporting seamless mobility and users' demands on multimedia applications imply that future WLAN handoff solutions must meet the following requirements:

- *Backward compatibility*: As a number of WLANs have been deployed using IEEE 802.11b/g, new handoff solutions must be compatible with the current legacy WLAN systems [118].

- *Application diversity*: Various applications will be proposed for WLANs. Future handoff schemes must be designed to meet users' demands for value-added services.

- *Integration with other heterogeneous networks*: To expand the territory of wireless services, new fast handoff solutions will be designed to provide broader radio coverage and seamless service for WLAN co-located with cellular networks, wireless LAN-based mesh networks, etc.

- *Network-controlled handoff*: The inherited IEEE 802.11 Standard requires handoffs to be managed, autonomously and independently, by each mobile station without pre-knowledge of the wireless environment. However, most current solutions pertain to *mobile station-controlled handoffs*. As wireless networks themselves are endowed with the capacity of leveraging considerable information regarding their topology and station proximity, *network-controlled handoff* schemes will tend to be significantly more present in future proposals.

- *Continuous signal monitoring capacity*: Since most cellular networks provide facilities with continuously monitoring signal quality between mobile stations

and all neighboring APs, future WLANs should develop these types of functionalities for mobile stations, using multi-mode radio interfaces.

- *Load balancing*: As overloaded APs cannot provide services to newly handoff mobile stations, new fast handoff scheme must introduce novel techniques to leverage the workload in wireless networks. Several load balancing solutions have been reported in the literature, such as cell-breathing [162] [163], yet further progress is required for future handoff designs.

- *Handoff in a mixed architecture*: Given that the inherited IEEE standard defines two architectural components, i.e. infrastructure and ad hoc mode; it is possible for mobile stations to operate with both modes simultaneously, using double radio interfaces. Thus, new fast handoff solutions will be valuable for handoffs in mixed architectures, such as ad hoc-assisted handoff scheme [164].

5.4 Proposed fast MAC layer handoff scheme

Our research objective is to provide fast handoff support for mobile hosts roaming with ongoing real-time applications in wireless LANs. Hence, the main research motivation consists of minimizing handoff delays and packet loss rates caused by handoff. We assume that mobile stations (or nodes) can completely skip the handoff detection phase using any triggers provided by the physical layer. Such fact is confirmed through the experimental study conducted in [120]. Additionally, the proposed fast MAC layer handoff scheme consists of fast scanning process, which is described as follows:

When the received signal strength is below a pre-defined threshold, the physical layer of a handoff mobile node (MN) sends a physical layer event notification (also called L1 trigger) to the MAC layer. However, it is a challenging issue to define this threshold as this depends on the real-life practical circumstance and affected by the surrounding interferences.

Upon receiving a trigger from the physical layer, an MN with ongoing real-time communication with a correspondent node (CN), launches a fast scanning procedure by analyzing its currently associated channel. In our algorithm, a channel analyzer module is defined and designed to store the current associated channel information.

Then, the MN switches to the next channel and achieves wireless medium access control using normal channel access procedure, e.g. carrier sense multiple access

with collision avoidance (CSMA/CA). The MN broadcasts a probe request on the examined channel and starts a probe timer. Then, it listens to the channel, waiting for probe responses sent by APs within range. If no response is received before the *MinChannelTime* is reached, the MN switches to next channel and performs an active scanning.

Once the first probe response is received, the MN immediately begins authentication with the AP that sent the response. This optimally minimizes the probe-waiting time on an examined channel. Upon successful authentication and reassociation, the MS completes the MAC layer (L2) handoff.

The advantage of the proposed approach is that probe delays can be reduced significantly, while only a subset of the allowed channels is scanned aside from the minimal probe-waiting time on each examined channel. As a result, handoff latencies and packet losses are reduced for mobile hosts roaming with real-time applications in progress. This proposal applies to fast movement cases as well as those where mobile stations need to handoff as quickly as possible. In addition, it requires neither AP modifications (such as SyncScan) nor pre-knowledge of the wireless network topology (such as the neighbor graphs approach and the selective scanning plus AP caching schemes). Moreover, the addition of a second radio interface for each mobile station is unnecessary (unlike MultiScan). Furthermore, simulation results for handoff latencies and packet losses are obtained from the same test bed (unlike channel masks and AP caching techniques), guaranteeing the consistency and credibility of results. However, this proposal also includes certain limitations, such as the possibility that the mobile does not select the best AP at the moment of handoff.

5.5 Performance evaluation

To evaluate the performance, simulations were conducted with SimulX [165], a C++ simulator developed at Louis-Pasteur University in France. Especially designed for IEEE 802.11 networks, SimulX also provides mobility support in IPv6 networks. The IEEE 802.11b standard [110] with 14 channels, mobile IPv6 (MIPv6) protocol [7] and the selective scanning plus AP caching schemes were already implemented. Based on these codes, we implement the IEEE 802.11b standard with 11 channels, The standard IEEE 802.11b with Min during which a mobile station only waits for *MinChannelTime* on each examined channel, the neighbor graphs schemes and our

proposed fast L2 handoff solution.

5.5.1 Network topology

The investigated scenario consists of a Mobile Node (MN) moving inside a building at an average speed of $1m/s$, communicating with a CN that sends UDP packets every $20ms$ to emulate $64kbps$ pulse code modulated voice stream packetized into 160 bytes. Figure 5.8 illustrates the simulation scenario.

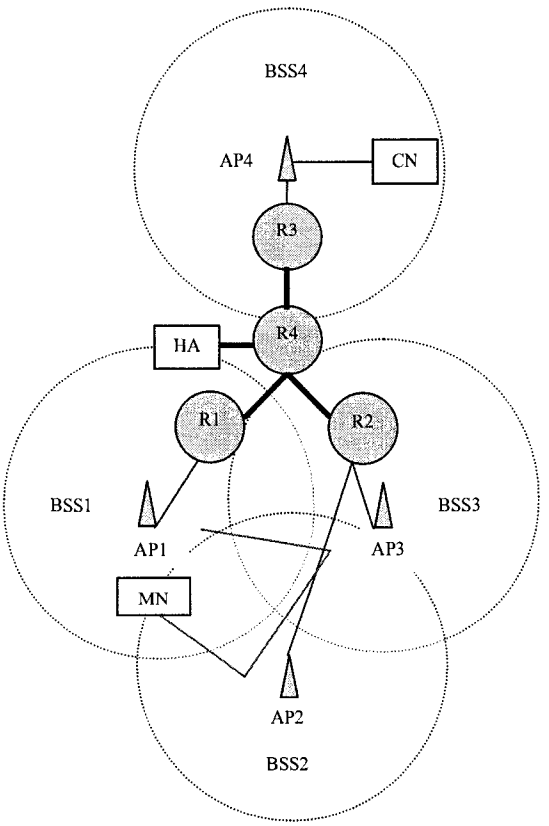

Figure 5.8 Network Topology Used for Simulation

The radio range of each network entity (including MN, CN and AP) equals $12m$. The $MinChannelTime$ is set to $17ms$ and the $MaxChannelTime$ is $38ms$ corresponding to Cisco devices [119]. AP_1 operates on the channel 1, AP_2 on channel 6, AP_3 on channel 11 and AP_4 on channel 6. The default beacon interval for each AP is

$100ms$. Five LANs with a $100Mbps$ capacity are present, and four WLANs of which the transmission rate ranging from 2 to $11Mbps$.

The MN performs three movements: from AP_1 to AP_2, then to AP_3, before returning to AP_1. However, the following performance analysis is based on the simulation results of the last two movements as the MN performs a full scan for Selective Scanning plus Caching schemes and constructs neighbor graphs and non-overlapping graphs for the Neighbor Graphs approach during the first movement. Thus, the performance evaluation represents a fair comparison, as the first movement of the MN is excluded from the analysis.

5.5.2 Simulation results

Probe delays versus AP's capacity

Figure 5.9 shows the relationship between probe delay and AP's capacity. Our proposed scheme outperforms the other four handoff solutions: the IEEE 802.11b standard, the standard IEEE 802.11b with Min, the selective scanning plus AP caching (Selective + Caching) and the neighbor graphs approaches. This is because our proposed scheme enables mobile stations to quickly terminate the scanning procedure once they finds an available AP to associate with. The average probe delay of the proposed scheme equals $35.70ms$, compared to $210.20ms$ for the IEEE 802.11b standard, the performance gain is 83.02%; compared to $189.51ms$ for the standard with $MinChannelTime$ (Min), the gain is 81.16%; compared to $55.51ms$ for the selective plus caching, the MN spends 35.69% less of probe delay; compared to $55.51ms$ for the Neighbor Graphs, the performance gain is 35.69%.

Authentication delays versus AP's capacity

Figure 5.10 shows authentication delay and AP's capacity. Authentication delay decreases rapidly as AP's capacity increases. Neighbor Graphs provides better performance amongst all solutions. The average authentication delay is $1.57ms$ for the proposed scheme; $1.34ms$ for the standard IEEE 802.11b; $1.32ms$ for the selective scanning plus AP caching schemes; $1.26ms$ for the standard 802.11b with $MinChannelTime$; $1.24ms$ for the neighbor graphs schemes. Although the proposed fast handoff scheme requires the most authentication delay, all the time differences are less than $0.4ms$ compared with other approaches. We explain this by the fact

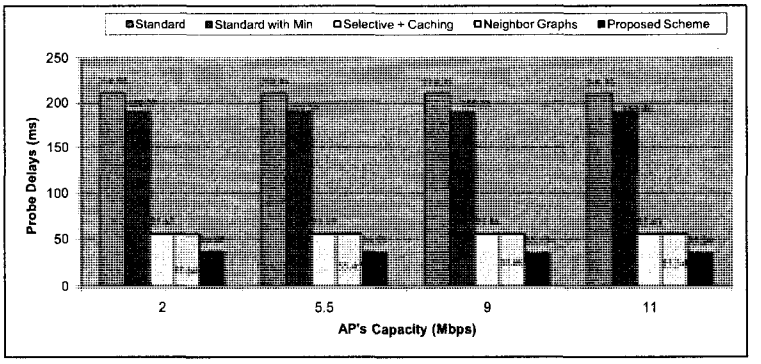

Figure 5.9 Probe Delays vs. AP's Capacity

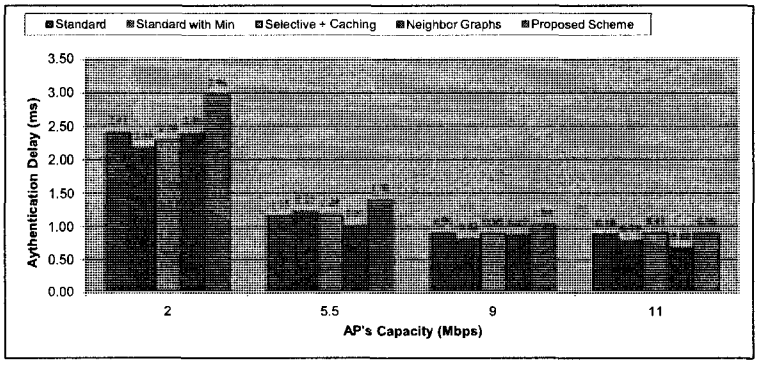

Figure 5.10 Authentication Delays versus AP's Capacity

that the processing time for executing the proposed scheme is a little bit longer than other solutions, since mobile nodes need to find their current associated channels at the moment of handoff.

Reassociation delays versus AP's capacity

Figure 5.11 shows the relationship between reassociation delay and AP's capacity. From the figure, we find that reassociation delay decreases rapidly as AP's capacity increases for both Standard with Min and the proposed scheme. Selective scanning plus AP Caching schemes deliver better performance amongst all the handoff solutions. The average reassociation delay is $1.63ms$; $1.65ms$ for the proposed fast handoff scheme and the standard IEEE 802.11b with Min (or $MinChannelTime$); $1.75ms$

for the neighbor graphs approaches; 1.80ms for the standard IEEE 802.11b.

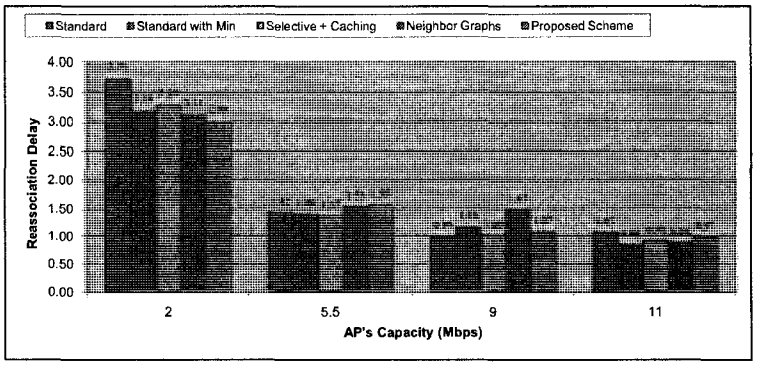

Figure 5.11 Reassociation Delays vs. AP's Capacity

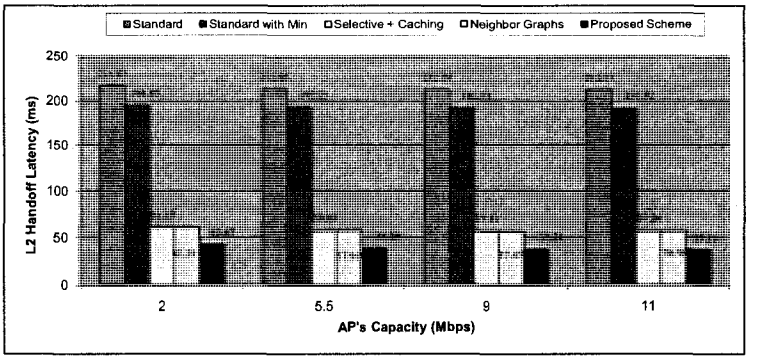

Figure 5.12 L2 Handoff Latencies versus AP's Capacity

L2 handoff latencies versus AP's capacity

Figure 5.12 illustrates the relationship between L2 handoff latency and AP's capacity. The increasing of AP's capacity leads to shorter L2 handover latencies. This is because the time taken for exchanging frames between the MS and the involved APs becomes lower due to higher transmission rate of the AP. Our proposed scheme delivers better performance than the other four handoff schemes. The average L2 handover delay of our proposal is 38.92ms, compared to 213.34ms for the IEEE 802.11b standard, the reduction is 81.76%; compared to 192.41ms for the standard with $MinChannelTime$

(or Min), the reduction is 79.77%; compared to 58.46ms for the selective scanning plus AP caching schemes, the reduction is 33.42%; compared to 58.38ms for the neighbor graphs approaches, the optimization is 33.33%.

L3 handoff delays with RO mode versus AP's capacity

Figure 5.13 illustrates the relationship between L3 handoff latency and AP's capacity for MIPv6 with Route Optimization (RO) Mode. The increasing of AP's capacity leads to shorter L3 handover latencies. We explain this by the fact that L2 handoff delays are an important component of L3 handoff latency. As a result, the lower the L2 handoff delay, the shorter L3 handoff latency. Our proposed scheme delivers better performance among all schemes. The average L3 handover delay is 69.91ms, compared to 273.33ms for the IEEE 802.11b standard, the performance gain is 74.42%; compared to 254.53ms for the standard with $MinChannelTime$ (or Min), the gain is 72.53%; compared to 125.27ms for the neighbor graphs approaches, the optimization is 44.19% compared to 123.45ms for the selective scanning plus AP caching schemes, the gain is 43.37%.

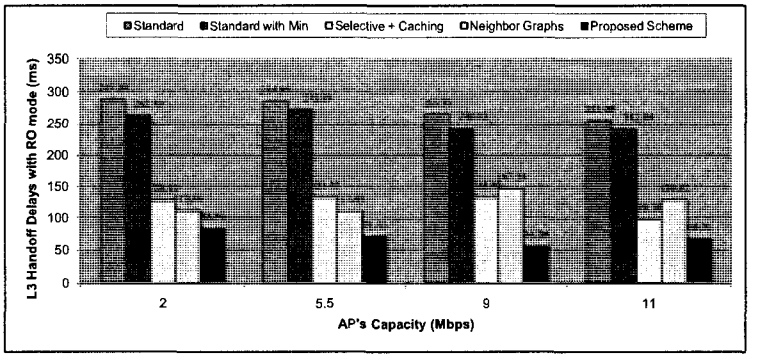

Figure 5.13 L3 Handoff Delays with RO Mode versus AP's Capacity

L3 handoff delays without RO mode versus AP's capacity

Figure 5.14 illustrates the relationship between L3 handoff latency and AP's capacity for MIPv6 without RO Mode. The increasing of AP's capacity leads to shorter L3 handover latencies. We explain this by the fact that L2 handoff delays are an important component of L3 handoff latency. As a result, the lower the L2 handoff delay, the

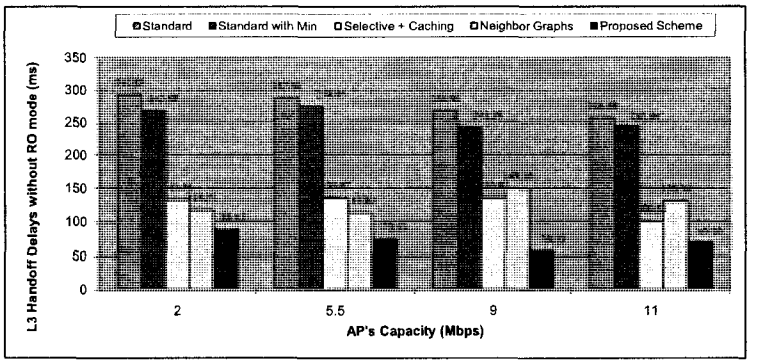

Figure 5.14 L3 Handoff Delays without RO Mode versus AP's Capacity

shorter L3 handoff latency. Our proposed scheme delivers better performance among all schemes. The average L3 handover delay is 72.42ms, compared to 276.02ms for the IEEE 802.11b standard, the performance gain is 73.76%; compared to 257.18ms for the standard with $MinChannelTime$ (or Min), the gain is 71.84%; compared to 127.90ms for the neighbor graphs approaches, the optimization is 43.38% compared to 126.33ms for the selective scanning plus AP caching schemes, the gain is 42.68%.

Packet loss rate versus AP's capacity

Figure 5.15 shows the relationship between packet loss rate and AP's capacity. Packet loss rate is defined as a ratio of the number of lost packets over the total number of transmitted packets at the application layer. Again, our proposed solution yields better performance than other schemes. The average packet loss rate for the proposed scheme equals 1.39%, compared to 2.55% for the IEEE 802.11b standard, the performance gain is 45.49%; compared to 2.29% for the standard with $MinChannelTime$ (Min), the optimization is 39.34%; compared to 1.94% for the selective scanning plus AP caching schemes, the reduction is 28.39%; compared to 1.72% for the neighbor graphs approach, the reduction is 19.57%. To maintain VoIP quality, the packet loss rate should be at or below 3% [166], thus our proposed fast handoff solution can meet this requirement.

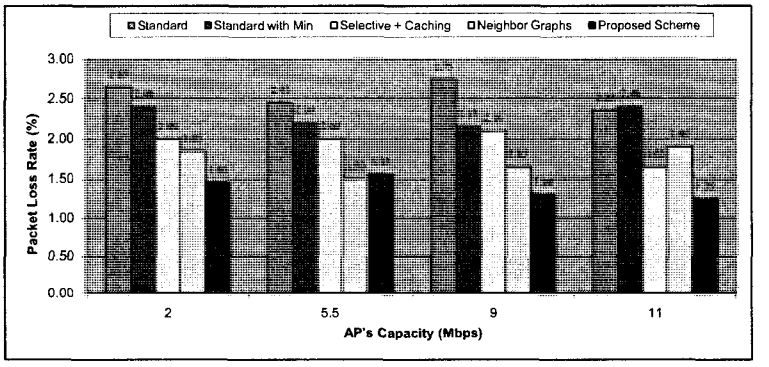

Figure 5.15 Packet Loss Rate versus AP's Capacity

5.6 Conclusion

This chapter first provides a survey of link-layer fast handoff schemes for mobile hosts roaming in WLANs. As shown above, providing seamless mobility and supporting real-time applications represent challenging issues for WLANs. To resolve this issue, this chapter also proposes a new MAC layer fast handoff scheme for mobile stations roaming in WLANs. Such scheme consists of minimizing the number of scanning channels and reducing the probe-waiting time on each examined channel to an optimal. To evaluate the efficiency of our proposal, we conduct simulations using the simulator SimulX. Simulation results show that our proposal deliver better performance than the handoff process of the standard IEEE 802.11b, the standard with *MinChannelTime* of which mobile stations only wait for *MinChannelTime* on each probed channel, selective scanning and AP caching techniques and the neighbor graphs approaches.

On the other hand, IEEE 802.11 Standard working groups are currently striving to enhance and standardize fast handoff management. Novel handoff approaches that meet the aforementioned design requirements (presented in Section 5.3) represent a promising field of research in the near future.

CHAPTER 6

CONCLUSION

This thesis proposes a new Access Router Tunneling Protocol (ARTP) to enable communication service provider to support seamless roaming for users coming from other service providers. Furthermore, a novel integrated architecture is designed for next-generation heterogeneous wireless networks. Based on the designed architecture, seamless handoff management schemes are proposed to facilitate IP layer (or layer three (L3)) mobility support with QoS provisioning. To evaluate the performance of the proposal, we employ analytical models to investigate the impact of various wireless system parameters on the handoff process. Moreover, simulations are carried out with OPNET Modeler v.12.0 to study the performance of seamless handoff procedures. In addition, since MAC layer (or layer two (L2)) handoff latency is an important component of L3 handoff delays, we also propose new fast MAC layer handoff schemes for MIPv6/WLANs environments in this thesis. And performance evaluation is done through simulations with the simulator SimulX. Major contributions of this research work are summarized in Section 5.1. Based on the results of this research, Section 5.2 identifies the limitations of this work and outlines the promising research areas within which future works may be expected.

6.1 Summary of contributions

This research makes four contributions to the state of the art in research in mobility management for next-generation heterogeneous wireless networks:

- Access Router Tunneling Protocol (ARTP)
- New integrated architecture for next-generation wireless networks
- Seamless handoff schemes for mobile IPv6-based wireless networks
- Fast MAC layer handoff scheme for MIPv6/WLANs environments

We summarize these four contributions individually in the following subsections.

6.1.1 Access router tunneling protocol

Access Router Tunneling Protocol (ARTP) enables mobile service providers to enlarge their business scope through offering communication services to subscribers from other operators. In addition, such protocol is used to establish bidirectional secure tunnels between adjacent radio access networks before actual handoff. Service Level Agreements (SLAs) are made between the involved operators. Within an SLA, certain QoS related parameters as well as security aspects are specified for each configured tunnel. As a result, ongoing multimedia sessions during handoff are guaranteed with QoS through pre-configured tunnels. This improves session quality and satisfies the mobile users' needs.

6.1.2 New integrated architecture

So far, disparate wireless networks have been deployed in the market, and these networks are expected to integrate with each other to provide ubiquitous and high data-rate services to roaming users. In this context, we propose a new integrated architecture to support fast and seamless roaming in next-generation wireless networks. The new architecture integrates different wireless communication systems using IPv6 as interconnection protocol. Besides, new network entities are introduced into the proposed architecture, such as eMAP and eHAAA.

6.1.3 Seamless handoff schemes

Classical mobility management protocols are not suitable for seamless roaming and QoS provisioning, we propose new seamless handoff schemes for mobile IPv6-based wireless networks. Such schemes consist of enabling mobile users to utilize their previous valid IP addresses in a new visiting network, and employ the pre-configured bidirectional tunnels during handoff. By doing so, new routing functionalities are added into access router. As a consequence, the new access router can provide routing services to mobile nodes that come from a trusted neighborhood. To evaluate the performance, we adopt two analytical models to investigate handoff related signaling costs, packet delivery costs and total costs, and analyze various wireless system parameters such as cell residence time, user velocity, cell size, domain size, session arrival rate, wireless link cost, session-to-mobility ratio on these costs. Additionally,

simulations are carried out with OPNET Modeler v.12.0. Both numerical results and simulation results show that the proposed seamless handoff schemes yield better performance than MIPv6 and its enhancements.

6.1.4 Fast MAC layer handoff scheme

IEEE 802.11-based Wireless Local Area Networks (WLANs) are experiencing rapid growth these days. However, the legacy standard cannot provide enough support for mobility management, in particular, handoff management. Therefore, we propose a fast handoff scheme in an MIPv6/WLANs environment for MAC layer mobility management. The new approach consists of minimizing the number of probed channels during handoff and reducing the probe-waiting time on each examined channel. Performance evaluation is carried out through simulations with the simulator SimulX. Of which, the results show that our proposal deliver better performance than the Standard IEEE 802.11b, the Standard IEEE 802.11b with *MinChannelTime*, and two well-documented solutions in the literature: Selective scanning plus AP caching techniques and the Neighbor Graphs mechanisms.

6.2 Limitations of the thesis

Even though the proposed SMIPv6 schemes deliver better performance than MIPv6 and its enhancements such as HMIPv6, FMIPv6, F-HMIPv6, we find that such schemes are always host-centric. In other words, they require mobile nodes to signal mobility management to their home agents and all active correspondent nodes. However, in case where a mobile node has no capability to transmit the mobility related signaling, host-centric mobility management protocols will be no longer functional. This is one of the limitations of this research work, as SMIPv6 schemes present a host-centric solution. In other words, the proposed seamless handoff schemes SMIPv6 require mobile nodes possessing the capability of sending binding update messages to their home agents or all active correspondent nodes. In case where the mobile has no such capability, SMIPv6 schemes cannot function properly. Under the circumstances, the interworking with proxy mobile IPv6 (PMIPv6) needs to be taken into consideration in the near future.

Since performance is evaluated through analytical models and simulations, the

utility of SMIPv6 protocol cannot be fully implemented. More specifically, simulations are conducted only for the case of predictive SMIPv6. We cannot compare the simulation results with those of HMIPv6, FMIPv6 and F-HMIPv6 since OPNET Modeler has not implemented these handoff management schemes. On the other hand, using analytical models, we analyze different wireless system parameters, such as user density, user velocity, domain size, session to mobility ratio on the handoff performance. The numerical results show that these parameters of wireless systems have an important impact on the system performance. But how to select the values of parameters that allow the SMIPv6 to reach the goal of total seamless is a challenging issue.

In addition, this thesis focuses on mobility management issue in next-generation wireless networks. However, in reality, each time before a mobile user obtains the services, it is necessary to undergo an authentication process. This results in longer and additional authentication delays during handoff. On the other hand, different wireless systems utilize different authentication protocol. The integration and interworking of these authentication mechanisms are a difficult issue. As a result, new fast authentication mechanisms are required along with seamless mobility management.

This thesis also proposes a novel access router tunneling protocol, which allows routers to pre-configure bidirectional secure tunnels before actual handoff. Moreover, such tunnels are utilized to guarantee certain quality of service for ongoing multimedia sessions during handoff. However, it is hard to implement this protocol as programmable routers are required to add new routing functionalities. This is one of the limitations of this research work. The difficulty of implementation also leads to the impossibility of evaluating the performance of this protocol.

This thesis also proposes a new integrated architecture for next-generation heterogeneous wireless networks. As seen, most of the proposed integrated architectures have exploited analytical models to evaluate the performance. However, as analytical models are always based on a number of assumptions, this makes the obtained numerical results questionable. Hence, implementation through setup real testbeds will be preferable in the near future.

6.3 Future work

Several open research issues are seeking for future research. Since new MIPv6-oriented enhancements are brought to the working items within the IETF working groups every day at every moment, we believe that new seamless handoff management protocols are required to not only reduce handoff latencies but also to minimize packet losses caused by the handoff process.

According to the limitations of SMIPv6, the integration and interworking between SMIPv6 and PMIPv6 will be taken into account in the near future. This prospective topic aims to resolve the problem when a mobile host cannot signal its mobility to the home agent or any peer node. PMIPv6 defines the way of using mobile access gateway to signal mobility management on behalf of the associated mobile nodes, it is possible for SMIPv6 to employ such access gateway for the purpose of signalization. In other words, the functionalities of mobile access gateway can be added into access router in the SMIPv6-based wireless systems.

Besides the seamless mobility management issue, new integrated authentication protocol is required to reduce the delays during handoff. This protocol should be combined with SMIPv6, together to guarantee session continuity and seamless handoff management.

As the implementation of access router tunneling protocol is a hard issue to resolve, in the near future, virtual access router (or node) can be developed in order to evaluate the performance of using this protocol. Such virtual access router should implement the tables such as Neighbor Table, Forwarding Tunnel Table, Reverse Tunnel Table, Tunneling Key Table, etc. In addition, new routing policies will be added into the virtual access router.

References

[1] 3G Americas, "The world wireless market-subscriptions by technology (Dec. 2007)," Internet: www.3gamericas.org/English/Statistics/q42007_1.cfm. [April 28, 2008].

[2] S.Y. Hui and K.H. Yeung, "Challenges in the migration to 4G mobile systems," *IEEE Communications Magazine*, vol. 41, no. 12, pp. 54-59, December 2003.

[3] I.F. Akyildiz, S. Mohanty and J. Xie, "A ubiquitous mobile communication architecture for next-generation heterogeneous wireless systems," *IEEE Communications Magazine*, vol. 43, no. 6, pp. 529-536, June 2005.

[4] The Internet Engineering Task Force (IETF). Internet: www.ietf.org. [April 23, 2008].

[5] C. Perkins, "IP mobility support for IPv4," *IETF RFC3344*, August 2002.

[6] C. Perkins, "IP mobility support for IPv4, revised," *IETF draft*, draft-ietf-mip4-rfc3344bis-06.txt (work in progress), March 2008.

[7] D. Johnson, C. Perkins and J. Arkko, "Mobility support in IPv6," *IETF RFC3775*, June 2004.

[8] H. Soliman, C. Castelluccia, K. El Malki and L. Bellier, "Hierarchical mobile IPv6 mobility management (HMIPv6)," *IETF RFC4140*, August 2005.

[9] H. Soliman, C. Castelluccia, K. El Malki and L. Bellier, "Hierarchical mobile IPv6 mobility management (HMIPv6)," *IETF draft*, draft-ietf-mipshop-4140bis-02.txt (work in progress), March 2008.

[10] R. Koodli, "Fast handovers for mobile IPv6," *IETF RFC4068*, July 2005.

[11] R. Koodli, "Mobile IPv6 fast handovers," *IETF draft*, draft-ietf-mipshop-fmipv6-rfc4068bis-07.txt (work in progress), April 2008.

[12] P. McCann, "Mobile IPv6 fast handovers for 802.11 networks," *IETF RFC4260*, November 2005.

[13] S. Ryu, J. Choi and Y. Mun, "Enhanced fast handover for mobile IPv6 based on IEEE 802.11 network," *IETF draft*, draft-mun-mipshop-efh-fast-mipv6-01.txt (work in progress), October 2005.

[14] B. Xia and X. Qin, "Flexible fast handover for mobile IPv6," *IETF draft*, draft-qin-mipshop-flexible-fmip-00.txt (work in progress), June 2007.

[15] H. Yokota and G. Dommety, "Mobile IPv6 fast handovers for 3G CDMA networks," *IETF draft*, draft-ietf-mipshop-3gfh-07.txt (work in progress), April 2008.

[16] H.Y. Jung and S.J. Koh, "Fast handover support in hierarchical mobile IPv6," in *Proceedings of the 6th International Conference on Advanced Communication Technology (ICACT 2004)*, 9-11 February 2004, pp. 551-554.

[17] H.Y. Jung, H. Soliman, S.J. Koh and J.Y. Lee, "Fast handover for hierarchical MIPv6 (F-HMIPv6)," *IETF draft*, draft-jung-mobopts-fhmipv6-00.txt (work in progress), April 2005.

[18] H.Y. Jung, H. Soliman, S.J. Koh and N. Takamiya, "Fast handover for hierarchical MIPv6 (F-HMIPv6)," *IETF draft*, draft-jung-mipshop-fhmipv6-00.txt (work in progress), October 2005.

[19] H.Y. Jung, E.A. Kim, J.W. Yi and H.H. Lee, "A scheme for supporting fast handover in hierarchical mobile IPv6 networks," *ETRI Journal*, vol. 27, no. 6, pp. 798-801, December 2005.

[20] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhury and B. Patil, "Proxy mobile IPv6," *IETF draft*, draft-ietf-netlmm-proxymip6-12.txt (work in progress), April 2008.

[21] V. Devarapalli, S. Gundavelli, K. Chowdhury and A. Muhanna, "Proxy mobile IPv6 and mobile IPv6 interworking," *IETF draft*, draft-devarapalli-netlmm-pmipv6-mipv6-01.txt (work in progress), April 2007.

[22] M. Sahasrabudhe and V. Devarapalli, "Proxy mobile IPv6 and mobile IPv4 interworking," *IETF draft*, draft-meghana-netlmm-pmipv6-mipv4-00.txt (work in progress), February 2008.

[23] M. Liebsch, L. Le and J. Abeille, "Route optimization for proxy mobile IPv6," *IETF draft*, draft-abeille-netlmm-proxymip6ro-01.txt (work in progress), November 2007.

[24] S. Jeon and Y. Kim, "Fast route optimization for PMIPv6 handover," *IETF draft*, draft-sijeon-netlmm-fastro-pmip6-00.txt (work in progress), February 2008.

[25] P.-S. Kim, S.-E. Kim, J.S. Jin and S.-C. Lee, "Proactive correspondent registration for proxy mobile IPv6 route optimization," *International Journal of Computer Science and Network Security*, vol. 7, no. 11, pp. 149-154, November 2007.

[26] C.M. Mueller and O. Blume, "Network-based mobility with proxy mobile IPv6," in *Proceedings of IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2007)*, 3-7 September 2007, pp. 1-5.

[27] S. Park, N. Kang and Y. Kim, "Localized proxy-MIPv6 with route optimization in IP-Based networks," *IEICE Transactions on Communications*, vol. E90-B, no. 12, pp. 3682-3686, December 2007.

[28] P. Kim, S. Kim and J. Jin, "Fast handovers for proxy mobile IPv6 without inter-MAG signaling," *IETF draft*, draft-pskim-netlmm-fastpmip6-00.txt (work in progress), November 2007.

[29] F. Xia and B. Sarikaya, "Mobile node agnostic fast handovers for proxy mobile IPv6," *IETF draft*, draft-xia-netlmm-fmip-mnagno-02.txt (work in progress), November 2007.

[30] S. Park, J.E. Lee, J. Choi and Y. Kim, "Fast localized proxy mobile IPv6 (FLP-MIPv6)," *IETF draft*, draft-park-netlmm-fastpmip-00.txt (work in progress), February 2007.

[31] I.F. Akyildiz, J. McNair, J.S.M. Ho, H. Uzunalioglu and W. Wang, "Mobility management in next-generation wireless systems," *Proceedings of the IEEE*, vol. 87, no. 8, pp. 1347-1384, August 1999.

[32] A. Quintero, O. Garcia and S. Pierre, "An alternative strategy for location update and paging in mobile networks," *Computer Communications*, vol. 27, no. 15, pp. 1509-1523, September 2004.

[33] H. Safa, S. Pierre and J. Conan, "An efficient location management scheme for PCS networks," *Computer Communications*, vol. 24, no. 14, pp. 1355-1369, August 2001.

[34] J. Manner and M. Kojo, "Mobility related terminology," *IETF RFC3753*, June 2004.

[35] E. Njedjou, P. Bertin and P. Reynolds, "Motivation for network controlled handoffs using IP mobility between heterogeneous wireless access networks," *IETF draft*, draft-njedjou-inter-an-handoffs-00.txt (work in progress), June 2003.

[36] L. Taylor, R. Titmuss and C. Lebre, "The challenges of seamless handover in future mobile multimedianetworks," *IEEE Personal Communications*, vol. 6, no. 2, pp. April 1999.

[37] M. Endler and V. Nagamuta, "General approaches for implementing seamless handover," in *Proceedings of the 2nd ACM international workshop on Principles of mobile computing (POMC'02)*, 30-31 October 2002, pp. 17-24.

[38] M. Liebsch, A. Singh, H. Chaskar, D. Funato and E. Shim, "Candidate access router discovery (CARD)," *IETF RFC4066*, July 2005.

[39] T.C. Schmidt and M. Waehlisch, "Seamless multicast handover in a hierarchical mobile IPv6 environment (M-HMIPv6)," *IETF draft*, draft-schmidt-waehlisch-mhmipv6-04.txt (work in progress), November 2005.

[40] J. Xie, "Mobility management in next-generation all-IP-based wireless systems," *Ph.D. Thesis*, Georgia Institute of Technology, April 2004.

[41] R. Fan, J. Li, H. Le and S. Cheng,"Performance analysis of TCP with the support of seamless handover," in *Proceedings of International Conference on Communication Technology (ICCT 2003)*, 9-11 April 2003, pp. 825-828.

[42] T. Narten, E. Nordmark, W. Simpson and H. Soliman, "Neighbor discovery for IP version 6 (IPv6)," *IETF RFC4861*, September 2007.

[43] S. Thomson, T. Narten and T. Jinmei, "IPv6 stateless address autoconfiguration," *IETF RFC4862*, September 2007.

[44] E. Fogelstroem, A. Jonsson and C. Perkins, "Mobile IPv4 regional registration," *IETF RFC4857*, June 2007.

[45] J. Loughney, "Seamless mobility concerns," *IETF draft*, draft-loughney-seamoby-concerns-00.txt (work in progress), November 2000.

[46] R. Droms, J. Bound, B. Volz, T. Lemon, C. Perkins and M. Carney, "Dynamic host configuration protocol for IPv6 (DHCPv6)," *IETF RFC3315*, July 2003.

[47] A. Conta and S. Deering, "Generic packet tunneling in IPv6 specification," *IETF RFC2473*, December 1998.

[48] S. Kent and R. Atkinson, "IP encapsulating security payload (ESP)," *IETF RFC2406*, November 1998.

[49] A. Conta, S. Deering and M. Gupta, "Internet control message protocol (ICMPv6) for the internet protocol version 6 (IPv6) specification," *IETF RFC4443*, March 2006.

[50] C. Kaufman, "Internet key exchange (IKEv2) protocol," *IETF RFC4306*, December 2005.

[51] E. Gustafsson and A. Jonsson, "Always best connected," *IEEE Wireless Communications*, vol. 10, no. 1, pp. 49-55, February 2003.

[52] I.F. Akyildiz, J. Xie and S. Mohanty, "A survey of mobility management in next-generation all-IP-based wireless systems," *IEEE Wireless Communications*, vol. 11, no. 4, pp. 16-28, August 2004.

[53] J.-C. Chen and T. Zhang, *IP-Based Next-Generation Wireless Networks: Systems, Architectures, and Protocols*. New York: Wiley-Interscience, 2004.

[54] S. Deering and R. Hinden, "Internet protocol, version 6 (IPv6) specification," *IETF RFC2460*, December 1998.

[55] J. Arkko, G. Kuijpers, H. Soliman, J. Loughney and J. Wiljakka, "Internet protocol version 6 (IPv6) for some second and third generation cellular hosts," *IETF RFC3316*, April 2003.

[56] L. Dimopoulou, G. Leoleis and I.S. Venieris, "Fast handover support in a WLAN environment: challenges and perspectives," *IEEE Network*, vol. 19, no. 3, pp. 14-20, May-June 2005.

[57] N. Moore, "Optimistic duplicate address detection (DAD) for IPv6," *IETF RFC4429*, April 2006.

[58] J. Arkko, V. Devarapalli and F. Dupont, "Using IPsec to protect mobile IPv6 signaling between mobile nodes and home agents," *IETF RFC3776*, June 2004.

[59] J. Kempf, J. Wood and G. Fu, "Fast mobile IPv6 handover packet loss performance: measurements for emulated real time traffic," in *Proceedings of IEEE Wireless Communications and Networking Conference (WCNC 2003)*, 16-20 March 2003, pp. 1230-1235.

[60] S. Haseeb and A.F. Ismail, "Handoff latency analysis of mobile IPv6 protocol variations," *Computer Communications*, vol. 30, no. 4, pp. 849-855, February 2007.

[61] 3GPP, "3GPP a global initiative." Internet: www.3gpp.org. [April 28, 2008].

[62] 3GPP2, "Third generation partnership project 2 (3GPP2)." Internet: www.3gpp2.org. [April 28, 2008].

[63] IEEE 802.21, "Media independent handover services." Internet: ieee802.org/21/. [April 28, 2008].

[64] IEEE 802.11u, "IEEE P802.11 - task group u - meeting update." Internet: www.ieee802.org/11/Reports/tgu_update.htm. [April 28, 2008].

[65] IEEE 802.11u, "IEEE P802.11 - task group r - meeting update." Internet: grouper.ieee.org/groups/802/11/Reports/tgr_update.htm. [April 28, 2008].

[66] Q. Zhang, C. Guo, Z. Guo and W. Zhu, "Efficient mobility management for vertical handoff between WWAN and WLAN," *IEEE Communications Magazine*, vol. 41, no.11, pp. 102-108, November 2003.

[67] N. Nasser, A. Hasswa and H. Hassanein, "Handoffs in fourth generation heterogeneous networks," *IEEE Communications Magazine*, vol. 44, no.10, pp. 96-103, October 2006.

[68] J. McNair and F. Zhu, "Vertical handoffs in fourth-generation multinetwork environments," *IEEE Wireless Communications*, vol. 11, no.3, pp. 8-15, June 2004.

[69] S. Mohanty, "A new architecture for 3G and WLAN integration and inter-system handover management," *Wireless Networks*, vol. 12, no. 6, pp. 733-745, November 2006.

[70] W. Song, W. Zhuang and A. Saleh, "Interworking of 3G cellular networks and wireless LANs," *International Journal of Wireless and Mobile Computing*, vol. 2, no. 4, pp. 237-247, January 2008.

[71] D. Kim and A. Ganz, "Architecture for 3G and 802.16 wireless networks integration with QoS support," in *Proceedings of 2nd International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks (QShine'05)*, 22-24 August 2005, pp. 28.

[72] D. Niyato and E. Hossain, "Integration of IEEE 802.11 WLANs with IEEE 802.16-based multihop infrastructure mesh/relay networks: A game-theoretic approach to radio resource management," *IEEE Network*, vol. 21, no. 3, pp. 6-14, May-June 2007.

[73] 3GPP TS, "Feasibility study on 3GPP system to wireless local area network (WLAN) interworking (release 7)," *3GPP TR 22.934 v7.0.0*, June 2007.

[74] 3GPP TS, "3GPP system to wireless local area network (WLAN) interworking; system description (release 7)," *3GPP TS 23.234 v7.6.0*, December 2007.

[75] 3GPP2 TSG-S, "3GPP2 - WLAN interworking - stage 1 requirements," *S.R0087-0 v1.0*, July 2004.

[76] 3GPP2 TSG-S, "cdma2000 – WLAN Interworking," *S.R0087-A v1.0*, February 2006.

[77] A.K. Salkintzis, C. Fors and R. Pazhyannur, "WLAN-GPRS integration for next-generation mobile data networks," *IEEE Wireless Communications*, vol. 9, no. 5, pp. 112-124, October 2002.

[78] A.K. Salkintzis, "Interworking techniques and architectures for WLAN/3G integration toward 4G mobile data networks," *IEEE Wireless Communications*, vol. 11, no. 3, pp. 50-61, June 2004.

[79] S.-L. Tsao and C.-C. Lin, "Design and evaluation of UMTS-WLAN interworking strategies," in *Proceedings of 2002 IEEE 56th Vehicular Technology Conference (VTC 2002-Fall)*, 24-28 September 2002, pp. 777-781.

[80] S.-L. Tsao and C.-C. Lin, "VGSN: a gateway approach to interconnect UMTS/WLAN networks," in *Proceedings of the 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2002)*, 15-18 September 2002, pp. 275-279.

[81] M.M. Buddhikot, G. Chandranmenon, S. Han, Y.-W. Lee, S. Miller and L. Salgarelli, "Design and implementation of a WLAN/cdma2000 interworking architecture," *IEEE Communications Magazine*, vol. 41, no. 11, pp. 90-100, November 2003.

[82] M. Jaseemuddin, "An architecture for integrating UMTS and 802.11 WLAN networks," in *Proceedings of the 8th IEEE International Symposium on Computers and Communication (ISCC 2003)*, 30 June-3 July 2003, pp. 716-723.

[83] J.-Y. Song, H.J. Lee, S.-H. Lee, S.-W. Lee and D.-H. Cho, "Hybrid coupling scheme for UMTS and wireless LAN interworking," *AEU - International Journal of Electronics and Communications*, vol. 61, no. 5, pp. 329-336, May 2007.

[84] K. Ahmavaara, H. Haverinen and R. Pichna, "Interworking architecture between 3GPP and WLAN systems," *IEEE Communications Magazine*, vol. 41, no. 11, pp. 74-81, November 2003.

[85] F.A. Phiri and M.B.R. Murthy, "WLAN-GPRS tight coupling based interworking architecture with vertical handoff support," *Wireless Personal Communications*, vol. 40, no. 2, pp. 137-144, January 2007.

[86] Q.-T. Nguyen-Vuong, L. Fiat and N. Agoulmine, "An architecture for UMTS-WIMAX interworking," in *Proceedings of the 1st International Workshop on Broadband Convergence Networks (BcN 2006)*, 7 April 2006, pp. 1-10.

[87] M. Buddhikot, G. Chandranmenon, S. Han, Y.W. Lee, S. Miller and L. Salgarelli, "Integration of 802.11 and third-generation wireless data networks," in *Proceedings of 22nd Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM 2003)*, 30 March-3 April 2003, pp. 503-512.

[88] D. Wisely and E. Mitjana, "Paving the road to systems beyond 3G - the IST BRAIN and MIND projects," *Journal of Communications and Networks*, vol. 4, no. 4, pp. 292-301, December 2002.

[89] C. Makaya and S. Pierre, "An architecture for seamless mobility support in IP-based next-generation wireless networks," *IEEE Transactions on Vehicular Technology*, accepted for publication, June 2007.

[90] P. Calhoun, J. Loughney, E. Guttman, G. Zorn and J. Arkko, "Diameter base protocol," *IETF RFC3588*, September 2003.

[91] J. Kempf, P. Calhoun, G. Dommety, S. Thalanany, A. Singh, P.J. McCann and T. Hiller, "Bidirectional edge tunnel handover for IPv6," *IETF draft*, draft-kempf-beth-ipv6-02.txt (work in progress), September 2001.

[92] Y. Gwon and A. Yegin, "Enhanced forwarding from the previous care-of address (EFWD) for fast handovers in mobile IPv6," in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC 2004)*, 21-25 March 2004, pp. 861-866.

[93] M.H. Habaebi, "Macro/micro-mobility fast handover in hierarchical mobile IPv6," *Computer Communications*, vol. 29, no. 5, pp. 611-617, March 2006.

[94] X. Pérez-Costa and H. Hartenstein, "A simulation study on the performance of Mobile IPv6 in a WLAN-based cellular network," *Computer Networks*, vol. 40, no. 1, pp. 191-204, September 2002.

[95] X. Pérez-Costa, M. Torrent-Moreno and H. Hartenstein, "A Performance comparison of mobile IPv6, hierarchical Mobile IPv6, fast handovers for mobile IPv6 and their combination," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 7, no. 4, pp. 5-19, October 2003.

[96] Y. Gwon, J. Kempf and A. Yegin, "Scalability and robustness analysis of mobile IPv6, fast mobile IPv6, hierarchical mobile IPv6, and hybrid IPv6 mobility protocols using a large-scale simulation," in *Proceedings of the 2004 IEEE International Conference on Communications (ICC 2004)*, 20-24 June 2004, pp. 4087-4091.

[97] I.F. Akyildiz and W. Wang, "A dynamic location management scheme for next-generation multitier PCS systems," *IEEE Transactions on Wireless Communications*, vol. 1, no. 1, pp. 178-189, January 2002.

[98] S. Pack and Y. Choi, "A study on performance of hierarchical mobile IPv6 in IP-based cellular networks," *IEICE Transactions on Communications*, vol. E87-B, no. 3, pp. 462-469, March 2004.

[99] L.J. Zhang and S. Pierre, "Performance analysis of fast handover for hierarchical MIPv6 in cellular networks," in *Proceedings of the 2008 IEEE 67th Vehicular Technology Conference (VTC2008-Spring)*, 11–14 May 2008.

[100] J.G. Markoulidakis, G.L. Lyberopoulos and M.E. Anagnostou, "Traffic model for third generation cellular mobile telecommunication systems," *Wireless Networks*, vol. 4, no. 5, pp. 389-400, August 1998.

[101] G. Wan and E. Lin, "Cost reduction in location management using semi-realtime movement information," *Wireless Networks*, vol. 5, no. 4, pp. 245-256, July 1999.

[102] H. Xie, S. Tabbane and D.J. Goodman, "Dynamic location area management and performance analysis," in *Proceedings of the 1993 IEEE 43rd Vehicular Technology Conference*, 18-20 May 1993, pp. 536-539.

[103] J.S.M. Ho and I.F. Akyildiz, "Mobile user location update and paging under delay constraints," *Wireless Networks*, vol. 1, no. 4, pp. 413-425, December 1995.

[104] S. Pack and Y. Choi, "Performance analysis of hierarchical mobile IPv6 in IP-based cellular networks," in *Proceedings of IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC 2003)*, 7-10 September 2003, pp. 2818-2822.

[105] Y.-B. Lin, "Reducing location update cost in a PCS network," *IEEE/ACM Transactions on Networking*, vol. 5, no. 1, pp. 25-33, February 1997.

[106] M. Woo, "Performance analysis of mobile IP regional registration," *IEICE Transactions on Communications*, vol. E86-B, no. 2, pp. 472-478, February 2003.

[107] X. Zhang, J.G. Castellanos and A.T. Campbell, "P-MIP: paging extensions for mobile IP," *Mobile Networks and Applications,* vol. 7, no. 2, pp. 127-141, April 2002.

[108] IEEE 802.11, "Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications," IEEE, 1999.

[109] IEEE 802.11a, "Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications: high-speed physical layer in the 5 GHz band," IEEE, 1999.

[110] IEEE 802.11b, "Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications: higher-speed physical layer extension in the 2.4 GHz," IEEE, September 1999.

[111] IEEE 802.11g, "Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications, amendment 4: further higher data rate extension in the 2.4 GHz band," IEEE, June 2003.

[112] IEEE 802.11i, "Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications, amendment 6: medium access control (MAC) security enhancements," IEEE, July 2004.

[113] M.S. Bargh, R.J. Hulsebosch, E.H. Eertink, A. Prasad, H. Wang and P. Schoo, "Fast authentication methods for handovers between IEEE 802.11 wireless LANs," in *Proceedings of the 2nd ACM International Workshop on Wireless Mobile Applications and Services on WLAN Hotspots (WMASH 2004)*, 1 October 2004, pp. 51-60.

[114] IEEE 802.11F, "IEEE trial-use recommended practice for multi-vendor access point interoperability via an inter-access point protocol across distribution systems supporting IEEE 802.11 operation," IEEE, July 2003.

[115] C. Rigney, S. Willens, A. Rubens and W. Simpson, "Remote authentication dial in user service (RADIUS)," *IETF RFC2865*, June 2000.

[116] P.-J. Huang, Y.-C. Tseng and K.-C. Tsai, "A fast handoff mechanism for IEEE 802.11 and IAPP networks," in *Proceedings of the 63rd IEEE Vehicular Technology Conference (VTC 2006-Spring)*, 7-10 May 2006, pp. 966-970.

[117] C.-T. Chou and K.G. Shin, "An enhanced inter-access point protocol for uniform intra and intersubnet handoffs," *IEEE Transactions on Mobile Computing*, vol. 4, no. 4, pp. 321-334, July/August 2005.

[118] S. Pack, J. Choi, T. Kwon and Y. Choi, "Fast handoff support in IEEE 802.11 wireless networks," *IEEE Communications Surveys and Tutorials*, vol. 9, no.1, pp. 2-12, January 2007.

[119] A. Mishra, M. Shin and W. Arbaugh, "An empirical analysis of the IEEE 802.11 MAC layer handoff process," *ACM SIGCOMM Computer Communication Review*, vol. 33, no. 2, pp. 93–102, April 2003.

[120] H. Velayos and G. Karlsson, "Techniques to reduce IEEE 802.11b MAC layer handover time," in *Proceedings of the 2004 IEEE International Conference on Communications (ICC 2004)*, 20-24 June 2004, pp. 3844–3848.

[121] K. Kwon and C. Lee, "A fast handoff algorithm using intelligent channel scan for IEEE 802.11 WLANs," in *Proceedings of the 6th International Conference on Advanced Communication Technology (ICACT 2004)*, 9-11 February 2004, pp. 46-50.

[122] I. Ramani and S. Savage, "SyncScan: practical fast handoff for 802.11 infrastructure networks," in *Proceedings of the IEEE 24th Annual Conference on Computer Communications (INFOCOM 2005)*, 13-17 March 2005. pp. 675-684.

[123] Y. Liao and L. Gao, "Practical schemes for smooth MAC layer handoff in 802.11 wireless networks," in *Proceedings of the 2006 International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM 2006)*, 26-29 June 2006, pp. 181-190.

[124] S. Shin, A.G. Forte, A.S. Rawat and H. Schulzrinne, "Reducing MAC layer handoff latency in IEEE 802.11 wireless LANs," in *Proceedings of the 2nd ACM International Workshop on Mobility Management and Wireless Access Protocols (MobiWac'04)*, 1 October 2004, pp. 19-26.

[125] M. Shin, A. Mishra and W.A. Arbaugh, "Improving the latency of 802.11 handoffs using neighbor graphs," in *Proceedings of the 2nd International Conference on Mobile Systems, Applications, and Services (MobiSys'04)*, 6-9 June 2004, pp. 70-83.

[126] H.-S. Kim, S.-H. Park, C.-S. Park, J.-W. Kim and S.-J. Ko, "Selective channel scanning for fast handoff in wireless LAN using neighbor graph," in *Proceedings of the 2004 International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC2004)*, 6-8 July 2004.

[127] J. Montavont and T. Noël, "IEEE 802.11 handovers assisted by GPS information," in *Proceedings of the IEEE 2nd International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob'2006)*, 19-21 June 2006, pp. 166-172.

[128] S. Waharte, K. Ritzenthaler and R. Boutaba, "Selective active scanning for fast handoff in WLAN using sensor networks," in *Proceedings of the IFIP TC6/WG6.8 Conference on Mobile and Wireless Communication Networks (MWCN 2004)*, 25-27 October 2004, pp. 59-70.

[129] N. Mustafa, W. Mahmood, A.A. Chaudhry and C.M. Ibrahim, "Pre-scanning and dynamic caching for fast handoff at MAC layer in IEEE 802.11 wireless LANs," in *Proceedings of the IEEE 2nd International Conference on Mobile Adhoc and Sensor Systems (MASS 2005)*, 7-10 November 2005.

[130] C.-S. Li, Y.-C. Tseng and H.-C. Chao, "A neighbor caching mechanism for handoff in IEEE 802.11 wireless networks," in *Proceedings of 2007 International Conference on Multimedia and Ubiquitous Engineering (MUE 2007)*, 26-28 April 2007, pp. 48-53.

[131] J. Montavont, N. Montavont and T. Noël, "Enhanced schemes for L2 handover in IEEE 802.11 networks and their evaluations," in *Proceedings of the IEEE 16th Annual International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC 2005)*, 11-14 September 2005, pp. 1429-1434.

[132] V. Brik, A. Mishra and S. Banerjee, "Eliminating handoff latencies in 802.11 WLANs using multiple radios: applications, experience, and evaluation," in *Pro-*

ceedings of the Internet Measurement Conference 2005 (IMC'05), 19-21 October 2005, pp. 299-304.

[133] M. Ohta, "Smooth handover over IEEE 802.11 wireless LAN," *IETF draft*, draft-ohta-smooth-handover-wlan-00.txt (work in progress), June 2002.

[134] K. Ramachandran, S. Rangarajan and J. C. Lin, "Make-before-break MAC layer handoff in 802.11 wireless networks," in *Proceedings of the 2006 IEEE International Conference on Communications (ICC 2006)*, 11-15 June 2006, pp. 4818-4823.

[135] S. Shin, A.G. Forte and H. Schulzrinne, "Seamless layer-2 handoff using two radios in IEEE 802.11 wireless networks," *Columbia University Technical Report CUCS-018-06*, New York, NY, April 2006.

[136] H.-S. Kim, S.-H. Park, C.-S. Park, J.-W. Kim and S.-J. Ko, "Fast handoff scheme for seamless multimedia service in wireless LAN," in *Proceedings of the 5th International IFIP-TC6 Networking Conference (Networking 2006)*, 15-19 May 2006, pp. 942-953.

[137] C.-S. Park, H.-S. Kim, S.-H. Park, K.-H. Jang and S.-J. Ko, "Fast handoff algorithm using access points with Dual RF modules," in *Proceedings of the 3rd European Conference on Universal Multiservice Networks (ECUMN 2004)*, 25-27 October 2004, pp. 20-28.

[138] T. Manodham and T. Miki, "A novel AP for improving the performance of wireless LANs supporting VoIP," *Journal of Networks*, vol. 1, no. 4, pp. 41-48, August 2006.

[139] H. Wu, K. Tan, Y. Zhang and Q. Zhang, "Proactive scan: fast handoff with smart triggers for 802.11 wireless LAN," in *Proceedings of the 26th Annual IEEE Conference on Computer Communications (INFOCOM 2007)*, 6-12 May 2007. pp. 749-757.

[140] T. Manodham, L. Loyola and T. Miki, "Pre-active scan phase for the latency handover time issues in IEEE 802.11 wireless LANs," in *Proceedings of the International Conference on Communication and Broadband Networking (ICBN 2004)*, 7-9 April 2004.

[141] N. Montavont and T. Noël, "Anticipated handover over IEEE 802.11 networks," in *Proceedings of the IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob'2005)*, 22-24 August 2005, pp. 64-71.

[142] N. Montavont and T. Noël, "Fast movement detection in IEEE 802.11 networks," *Wireless Communications and Mobile Computing*, vol. 6, no.5, pp. 651-671, July 2006.

[143] V. Mhatre and K. Papagiannaki, "Using smart triggers for improved user performance in 802.11 wireless networks," in *Proceedings of the 4th International Conference on Mobile Systems, Applications and Services (MobiSys'06)*, 19-22 June 2006, pp. 246-259.

[144] C.-C. Tseng, K.-H. Chi, M.-D. Hsieh and H.-H. Chang, "Location-based fast handoff for 802.11 networks," *IEEE Communications Letters*, vol. 9, no. 4, pp. 304-306, April 2005.

[145] B. Park, Y.-H. Han and H. Latchman, "EAP: new fast handover scheme based on enhanced access point in mobile IPv6 networks," *International Journal of Computer Science and Network Security*, vol. 6, no. 9, pp. 69-75, September 2006.

[146] N. Jordan, A. Poropatich and R. Fleck, "Link-layer support for fast mobile IPv6 handover in wireless LAN based networks," in *Proceedings of the 13th IEEE Workshop on Local and Metropolitan Area Networks (LANMAN 2004)*, 25-28 April 2004, pp. 139-143.

[147] I. Samprakou, C. Bouras and T. Karoubalis, "Fast IP handoff support for VoIP and multimedia applications in 802.11 WLANs," in *Proceedings of the 6th IEEE International Symposium on a World of Wireless Mobile and Multimedia Networks (WoWMoM 2005)*, 13-16 June 2005, pp. 332-337.

[148] I. Samprakou, C. J. Bouras and T. Karoubalis, "Improvements on 'IP—IAPP' : a fast IP handoff protocol for IEEE 802.11 wireless and mobile clients," *Wireless Networks (WINET) Journal*, Special Issue on Broadband Wireless Multimedia, vol. 13, no. 4, pp. 497-510, August 2007.

[149] C.-C. Tseng, L.-H. Yen, H.-H. Chang and K.-C. Hsu, "Topology-aided cross-layer fast handoff designs for IEEE 802.11/mobile IP environments," *IEEE Communications Magazine*, vol. 43, no. 12, pp. 156-163, December 2005.

[150] A. Mishra, M. Shin and W.A. Arbaugh,"Context caching using neighbor graphs for fast handoffs in a wireless network," in *Proceedings of the IEEE 23rd Conference on Computer Communications, (INFOCOM 2004)*, 7-11 March 2004, pp. 351-361.

[151] A. Mishra, M. Shin, N. Petroni, T.C. Clancy and W.A. Arbaugh, "Proactive key distribution using neighbor graphs," *IEEE Wireless Communications*, vol. 11, no. 1, pp. 26-36, February 2004.

[152] J.-C. Chen, M.-C. Jiang and Y.-W. Liu, "Wireless LAN security and IEEE 802.11i," *IEEE Wireless Communications*, vol. 12, no. 1, pp. 27-36, February 2005.

[153] IEEE 802.1X, "IEEE standards for local and metropolitan area networks, port-based network access control," IEEE, December 2004.

[154] S. Pack, H. Jung, T. Kwon and Y. Choi, "SNC: a selective neighbor caching scheme for fast handoff in IEEE 802.11 wireless networks," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 9, no. 4, pp. 39-49, October 2005.

[155] S. Pack and Y. Choi, "Fast handoff scheme based on mobility prediction in public wireless LAN systems," *IEE Proceedings Communications*, vol. 151, no. 5, pp. 489-495, October 2004.

[156] S. Pack and Y. Choi, "Fast inter-AP handoff using predictive authentication scheme in a public wireless LAN," in *Proceedings of the IEEE Joint International Conference on Wireless LANs and Home Networks and Networking (Networks 2002)*, 26-29 August 2002, pp. 15-26.

[157] S. Pack and Y. Choi, "Pre-authenticated fast handoff in a public wireless LAN based on IEEE 802.1x model," in *Proceedings of the IFIP TC6/WG6.8 Working Conference on Personal Wireless Communications (PWC 2002)*, 23-25 October 2002, pp. 175-182.

[158] IEEE 802.11e, "Part 11: wireless LAN medium access control (MAC) and physical layer (PHY) specifications: amendment 8: medium access control (MAC) quality of service enhancements," IEEE, November 2005.

[159] D. Simone, "802.11k makes WLANs measure up," *Network World*, March 2004.

[160] P. Calhoun and B. O'Hara, "802.11r strengthens wireless voice," *Network World*, August 2005.

[161] F. Cacace and L. Vollero, "Managing mobility and adaptation in upcoming 802.21-enabled devices," in *Proceedings of the 4th ACM International Workshop on Wireless Mobile Applications and Services on WLAN Hotspots (WMASH 2006)*, 29 September 2006, pp. 1-10.

[162] Y. Bejerano and S.-J. Han, "Cell breathing techniques for load balancing in wireless LANs," in *Proceedings of the IEEE 25th International Conference on Computer Communications (INFOCOM 2006)*, 23-29 April 2006, pp. 1-13.

[163] H. Velayos, V. Aleo and G. Karlsson, "Load balancing in overlapping wireless LAN cells," in *Proceedings of the 2004 IEEE International Conference on Communications (ICC 2004)*, 20-24 June 2004, pp. 3833-3836.

[164] M. He, T.D. Todd, D. Zhao and V. Kezys, "Ad hoc assisted handoff for real-time voice in IEEE 802.11 infrastructure WLANs," in *Proceedings of the IEEE Wireless Communications and Networking Conference (2004 WCNC)*, 21-25 March 2004, pp. 201-206.

[165] SimulX Wireless Network Simulator Wiki Pages. [Internet]: wikinet.u-strasbg.fr/ simulx/index.php/HomePage, March 21, 2005 [May 5, 2008].

[166] A. Schmitter, A.Th. Schwarzbacher and T.D. Smith, "Analysis of network conformity with voice over IP specifications," in *Proceedings of the Irish Systems and Signals Conference (ISSC 2003)*, 1-2 July 2003, pp. 82-86.

[167] S. Pack and Y. Choi, "Performance analysis of fast handover in mobile IPv6 networks," in *Proceedings of the IFIP-TC6 8th International Conference on Personal Wireless Communications (PWC 2003)*, 23-25 September 2003, pp. 679-691.

[168] C. Makaya and S. Pierre, "An architecture for seamless mobility support in IP-based next-generation wireless networks," *IEEE Transactions on Vehicular Technology*, vol. 57, no. 2, pp. 1209-1225, March 2008.

[169] J. Hassan, H. Sirisena and B. Landfeldt, "Trust-based fast authentication for multiowner wireless networks," *IEEE Transactions on Mobile Computing*, vol. 7, no. 2, pp. 247-261, February 2008.

[170] P. Pongpaibool, P. Sotthivirat, S.I. Kitisin and C. Srisathapornphat, "Fast duplicate address detection for mobile IPv6," in *Proceedings of 15th IEEE International Conference on Networks (ICON 2007)*, November 2007, pp. 224-229.

APPENDIX

List of Publications

At the end of this thesis, we provide a list of publication within this research.

Patents pending [3]

- Li Jun Zhang and Laurent Marchand, "Handover Enabler," Patent application, US 11/410,206, April 25, 2006.
- Li Jun Zhang and Laurent Marchand, "Tunnel Establishment," Patent application, US 11/410,205, April 25, 2006.
- Li Jun Zhang, Laurent Marchand and Samuel Pierre, "Optimized Seamless Handover in Mobile IPv6 Networks," US Patent, US 60/674,356, April 25, 2005.

Book chapters [2]

- Li Jun Zhang and Samuel Pierre, "Fast MAC Layer Handoff Schemes in WLANs," in *Unlicensed Mobile Access Technology: Protocols, Architectures, Security, Standards and Applications*, Yan Zhang, Laurence T. Yang, Jianhua Ma (Ed.), Auerbach Publications, Taylor & Francis Group, USA, February, 2008.
- Li Jun Zhang and Samuel Pierre, "A Survey of Mobility Management Protocols in Next-Generation Wireless Networks," Submitted for publication, November 30, 2007.

Accepted journal papers [2]

- Li Jun Zhang and Samuel Pierre, "Evaluating the Performance of Fast Handover for Hierarchical MIPv6 in Cellular Networks," *Journal of Networks*, vol. 3, no. 6, pp. 36-43, June 2008.
- Li Jun Zhang and Samuel Pierre, "Optimizing the Performance of Handoff Management in Wireless LANs", International Journal of Computer Science and Network Security, accepted for publication, May 26, 2008..

Submitted journal papers [3]

- Li Jun Zhang and Samuel Pierre, "New IP Layer Seamless Handoff Schemes for Mobile IPv6 Networks," April 18, 2008.
- Li Jun Zhang and Samuel Pierre, "Optimized Seamless Handoff Schemes for Wireless and Mobile IPv6-Based Networks," April 7, 2008.
- Li Jun Zhang and Samuel Pierre, "Evaluating the Performance of Location Management for Mobile IPv6 and Its Enhancements," February 5, 2008.

Conference papers [7]

- Li Jun Zhang and Samuel Pierre, "Performance Analysis of Fast Handover for Hierarchical MIPv6 in Cellular Networks," in Proceedings of IEEE 67th Vehicular Technology Conference (VTC2008-Spring), Marina Bay, Singapore, May 11-14, 2008, pp. 2374-2378.
- Li Jun Zhang and Samuel Pierre, "Performance Enhancement for Mobility Management in Wireless LANs," in Proceedings of IEEE Wireless Communications & Networking Conference (WCNC 2008), Las Vegas, USA, March 31-April 3, 2008, pp. 1757-1762.
- Li Jun Zhang, Samuel Pierre and Liyan Zhang, "Fast Handoff Scheme with Accelerated Probe Function in Wireless LANs," in Proceedings of the 7th IASTED International Conferences on Wireless and Optical Communications (WOC 2007), Montreal, Canada, May 30-June 1, 2007, pp. 297-300.
- Li Jun Zhang, Samuel Pierre and Laurent Marchand, "A New Seamless Method to Support CDMA2000/WLAN Vertical Handover," in Proceedings of IEEE Canadian Conference on Electrical and Computer Engineering (CCECE 2006), Ottawa, Canada, May 7-10, 2006, pp. 214-218.
- Li Jun Zhang, Samuel Pierre and Laurent Marchand, "Optimization of Handover Performance for FMIPv6," in Proceedings of the IFIP International Conference on Intelligence in Communication Systems (INTELLCOMM 2005), Montreal, Canada, October 17-19, 2005, pp. 169-178.

- Li Jun Zhang, Samuel Pierre and Laurent Marchand, "Anticipated Seamless Handoff Scheme in Wireless LAN," CORS/Optimization Days 2006, Montreal, Canada, May 8-10, 2006.

- Li Jun Zhang and Samuel Pierre, "A Novel Architecture to Support Fast Authentication and Seamless Roaming in Next-Generation Heterogeneous Wireless Networks," submitted for publication, March 31, 2008.